MICROBIOLOGICAL REVIEWS, Sept. 1991, p. 371–394
0146-0749/91/030371-24$02.00/0

Vol. 55, No. 3

Control Site Location and Transcriptional Regulation in *Escherichia coli*

JULIO COLLADO-VIDES,[1] BORIS MAGASANIK,[1] AND JAY D. GRALLA[2]*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,[1] and Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90024-1569[2]

INTRODUCTION

Regulation of gene expression is one of the central themes of investigation in modern biology. The explosion of information from the study of gene regulation in eukaryotic organisms is becoming difficult to evaluate on its own and relies increasingly on information generated from studies of simpler prokaryotes. Certainly, one of the advantages of studies with the bacterium *Escherichia coli* is that it is one of the best, if not the best, known free-living biological system. We now have a reasonable level of understanding of a large number of individual systems of regulation. Particularly, the number of sigma 70 promoters studied may well represent the largest homogeneous body of knowledge of gene regulation at present. In this article we have collected and analyzed as a whole most of the sigma 70 promoters of *E. coli* and *Salmonella typhimurium* in which the regulatory sites are well defined, as well as all the known sigma 54 promoters of *E. coli* and of *Klebsiella pneumoniae*. This data set is now sufficiently large to allow the general principles of the circuitry of regulation to begin to emerge.

One of the most fundamental questions one can ask concerning control circuitry is why regulatory DNA elements are organized in particular arrangements. One presumes that these arrangements have been shuffled somewhat during the course of evolution to yield organizations that are appropriate to the specific requirements of individual regulatory systems. Using the very large data set now available, we have organized these data with the point of reference being the relative position of regulatory sites with respect to the site of initiation of transcription. We have included only those cases in which the regulatory sites have established functions and have arbitrarily excluded bacteriophage promoters, which have very special requirements as a class. Genes controlled by the factor sigma 54 are compiled separately since their primary distinction is that the positioning of their regulatory sites is very flexible.

Thus, there are two unique aspects of this review. One is the emphasis on the importance of the location of regulatory sites. The second is the completeness of the compilation, made possible by recent advances in the field of bacterial gene regulation. Taken together, these aspects allow certain inferences to be drawn about the range of regulatory mechanisms at the level of initiation of transcription. In addition, a catalog of operons with known regulatory sites has been assembled to provide a base for future analysis.

In an attempt to make sense of arrangements that might be associated with particular classes of promoters, a classification scheme is necessary. We will distinguish between two classes of locations of regulatory sites that affect promoters: proximal and remote. The boundaries of what is considered proximal will be set by comparison with the *lac* operon. Proximal sites will be defined as those that are between the upstream boundary of the *lac* cyclic AMP (cAMP) receptor protein (CRP) site and the downstream boundary of the RNA polymerase bound at the *lac* promoter. This means that sites that overlap approximately −65 and +20, the transcription start site being designated as +1, will be considered proximal. The implications of this definition will be discussed further below. All other regulatory sites will be defined as remote.

Individual promoters will be grouped into regulatory systems according to whether they are controlled by a particular regulator, irrespective of the number of binding sites in the DNA. Thus, for example, in this scheme the *lac* promoter belongs to two regulatory systems, one responsive to *lac* repressor and the other to the CRP. A promoter is called

* Corresponding author.

multiple if its regulatory region is connected to that of another promoter and both are controlled by at least one common element or transcribe the same gene. These would include examples such as *gal*, which has two overlapping promoters, or *malE,K* with its two divergent promoters. A promoter is termed complex if it is either multiple, like *gal*, or part of more than one regulatory system, like *lac*. All other promoters will be termed simple, i.e., isolated promoters controlled by only one regulatory system.

CATALOG OF SIGMA 70 PROMOTERS

Using the Medline data base, we have obtained information on the regulation of 119 promoters. The search was extended to include journal articles appearing up to October 1990. Of the 119 examples, 107 are sigma 70 promoters and 12 are sigma 54 promoters. The collection of sigma 70 promoters, in alphabetical order of regulators, is presented in Fig. 1. All the promoters are aligned with respect to the point of initiation of transcription. All the regulatory sites affecting one promoter are grouped in one line. Regions of multiple promoters have the individual promoters represented separately, one after the other. After presentation and a brief description of this catalog, the analysis of the organization of repression and activation systems will be considered sequentially.

Tables 1 and 2 display the entire data set of sigma 70 promoters and indicate their position in the classification scheme. Each promoter is categorized with respect to whether it is multiple, is subject to multiple systems of regulation, or is activatable or repressible and whether the regulatory sites are duplicated. Along with each promoter is listed one selected literature reference and the type of evidence implicating certain sites as regulatory elements. Various properties of the whole data set will now be discussed.

Of the 107 sigma 70 promoters, 49 (almost half) are complex promoters. Twenty-seven of these are complex because they are multiple promoters; 10 are complex because they are subject to multiple systems of regulation; and 12 are complex because of both reasons. This analysis could conceivably underestimate the number of complex promoters since new regulatory features could eventually be discovered. We will now consider briefly the subclasses of complex promoters, those that are multiple and those that are subject to multiple systems of regulation.

Nearly 40% of the cataloged promoters are defined as multiple. The obvious advantage to having multiple promoters is that each one of the set can be regulated differently, giving a potential for greater flexibility. Among the interdigitated promoters listed in Table 3 are those subject to different systems of regulation, those containing one constitutive and one regulated promoter, and examples of promoters that are recognized by different forms of polymerase. All the divergent promoter pairs share at least one regulatory site. Sharing regulatory sites in multiple promoters allows for differential repression since, as discussed below, the extent of repression can vary with the position of the operator.

Promoters subject to multiple systems of regulation make up one-fifth of the data base (although some of these are also multiple promoters). These isolated statistics indicating 37% multiple and 20% multiply regulated promoters leaves the impression that over 40% of promoters are simply regulated, suggesting that most promoters are regulated in isolation from most other promoters. The statistic is, however, rather misleading because many of these simple promoters are part of large regulons. Thus, although the seven LexA promoters are each defined as simple, they are all coregulated by the LexA repressor. If one takes account of connectivity via common regulatory molecules, only 4 of the 107 promoters appear to be regulated in isolation from other promoters. This very small number should decrease as more promoters are characterized, since some will be found to be coregulated with those included in this collection or with not yet described promoters.

NEGATIVE REGULATION OF SIGMA 70 PROMOTERS

Figure 1 contains 107 promoters, the large majority of which are repressible promoters. Of the 76 repressible promoters, 70 contain at least one operator in a proximal position. Of the remaining six, four have remote operators that overlap with a proximal activator site. The exceptions (those without apparent proximal regulatory elements) will be discussed below. This simple analysis of the data base indicates that almost all known promoters that contain operators have at least one that either is proximal or, in a very few cases, overlaps other elements for the proximal transcription machinery.

Such an arrangement of the negative regulatory apparatus has been proposed to reflect a requirement for bound repressors to touch the proximal transcription apparatus directly (42). Thus, in almost all cases, the repressor is close enough to touch either the polymerase or an activator which in turn is close enough to touch the polymerase. The specific mechanism for repression could vary considerably, but the involvement of a site that locates at least one repressor where it could touch polymerase or an activator seems a nearly universal arrangement in the negative regulation of sigma 70 promoters.

One area of potential uncertainty is the appropriateness of the definition of a proximal site. The polymerase itself protects the DNA strongly between positions −45 and +20 but probably occupies a space beyond these limits (44). The definition of proximal goes beyond this upstream limit to encompass the CRP site location in the *lac* promoter. It is known that CRP binds RNA polymerase in solution (105) and that its activation drops dramatically as its binding site is moved further upstream (40). Thus, it is plausible that polymerase can be touched from this CRP position (121) and that this ability weakens significantly as sites are moved out of the proximal region as defined here. In fact, this is an issue in only a few cases where the operators reside proximally but upstream from −45; these will be discussed under Variation in Operator Position, below.

There are 3 apparent exceptions to this arrangement in the 76 examples in this compilation: the *purR*, *nrd*, and *aroP* promoters (see PurR, Z, and TyrR regulators in Tables 1 and 2). The *purR* promoter is apparently controlled by only two remote operators, located around +100 and +200 (89, 124). The effect of these operators is very weak, only about twofold. Interference with elongation rather than initiation does not provide a strong repression, as has been observed in *lac* (36). Repression here may thus occur from downstream by interfering with transcription elongation. This may therefore be more an example of an elongation block than of promoter control and would not be considered an exception with regard to repression of transcription initiation. The repression site in the *aroP* promoter is in a downstream position bordering +30, only 10 bp removed from the border protected by the bound polymerase (21). The border of

protection by the bound TyrR protein is not known and could possibly approach the proximal +20 position more closely. If the polymerase and TyrR repressor each extends only 0.6 or 0.7 nm beyond the +20 and +30 sites, respectively, they would touch. Alternatively, it is possible either that this is a true but marginal exception or that the site really serves to block elongation, as occurs in *lac* and perhaps *purR*.

In the *nrd* promoter (156, 157), there are inverted and direct DNA repeats that extend from the remote upstream site to at least −57 within the proximal region. It has been suggested that some of these sites may be involved in regulation, but the genetics of the system is not yet well defined; because of this uncertainty in the location of regulatory sites, the *nrd* system will not be considered further in the analysis. This discussion indicates that of 107 promoters, these 3 without obvious proximal elements could conceivably be more apparent as exceptions than real.

The function of a repressor is to prevent the initiation of transcription, and this analysis indicates that in most or all cases, operator arrangements allow repression to occur by direct interference with the polymerase from a proximal site. This requirement for a proximal operator appears not to restrict the range of repression accessible to the sigma 70 transcription machinery since repressibility is known to vary over 3 orders of magnitude. Even within regulons, repressibility can vary considerably. These variations are apparently achieved by varying either the affinity of the operator site for the repressor or the specific position of the operator relative to the promoter.

In the following section we will analyze the role of the position of the proximal operator in the mechanism of repression. Then we will discuss the role played by duplication of operators as another way to increase the flexibility in the repression apparatus used by sigma 70 promoters.

Variation in Operator Position and the Mechanism of Repression

Analysis of the data base shows that when an operator type appears in more than one promoter (presumably in a regulon), it rarely appears in the same proximal position. For example, the seven TyrR-regulated promoters, eight of the nine LexA-regulated promoters, six of the nine PurR-regulated promoters, the three TrpR-regulated promoters, two of the five MetJ-regulated promoters, and at least two of the six ArgR-regulated promoters have operator sequences located in different proximal positions. As discussed below, this variation is in strong contrast to the relatively fixed positions of activator sites.

A plausible explanation for this extensive variation is that it allows each promoter to be repressed by the same protein but in a different manner. This receives credence from studies such as those in which the *lac* operator was moved to different proximal positions (68). The degree of repression varied considerably with the proximal position. Of the three operator positions studied, immediately upstream of the −35 domain, the spacer region, and downstream of the −10 domain overlapping the transcription initiation, the spacer confers highest repression. A repressor bound to the upstream position was rationalized to be less effective in preventing polymerase binding since the critical promoter −10 and initiation sites remain largely exposed, allowing polymerase substantial access to its binding site. A repressor bound over the −10 elements and start site occludes the most critical initiation region but suffers in competition with polymerase since polymerase can potentially form precursors to productive complexes by using the −35 element. By contrast, when the operator is in the spacer region, the repressor molecule can potentially occlude both elements and even begin binding before any initiating polymerases fully clear the promoter; after initiation the polymerase clears the spacer before it clears the −10 region and initiation site. In this model the position of the operator influences repressibility by determining when it is cleared and available for repressor binding and also by determining how effectively the bound repressor can occlude the determinants of polymerase binding.

These properties rationalize the effects of dramatically different positions, but even small changes could conceivably have important consequences. It is known that repression complexes may also contain RNA polymerase bound to DNA in an inactivate state (147), as implied in the above discussion. Conversely, it is also suspected that simple binding of a protein to an operator is not always sufficient for repression (46, 85). Repression thus appears in at least some cases to involve a potential interaction between the repressor and the RNA polymerase. The relative positions of the polymerase and the repressor will influence how they interact and potentially have consequences for the effectiveness of repression. The variations in repressibility resulting from variations in position could be due in part to repressors touching polymerase differently from the different locations.

This variability appears to be used by most of the regulons in the data set and may explain why regulon organizations may be preferable to large operon arrangements in many cases. Recall that there was tremendous diversity of operator positions within regulons (see above), such as the nine LexA-controlled genes displaying eight different operator positions. This regulon specifies products that are needed at different times and in different amounts during the SOS response. It might be difficult to construct a single large operon producing all of these proteins that would accommodate these needs. By separating the transcription units and placing similar operators in different positions, one achieves the potential for greater flexibility. Similar arguments can be made, for example, concerning the many promoters involved in aromatic amino acid metabolism. In these cases the differing positions of the *tyr* and *trp* operators might allow a more flexible response to physiological needs in this complex branched pathway.

The analysis of the effectiveness of repression requires that at least one operator be in a proximal position, where it can touch either the polymerase or the proximal polymerase recognition elements. This is based on the observation that virtually all repressible promoters have proximal operators, as discussed above. However, the definition of proximal position was based on the activator CRP, not on a repressor. The very large majority of operators overlap the actual polymerase footprint site from −45 to +20; however, in a few cases the operators are defined here as proximal but lie upstream from −45 or are slightly remote but coexist with a proximal activator site. In these cases it is possible that other mechanisms of repression are applicable, as will now be considered.

These examples with single operators upstream of −45 are the promoters of the CytR regulon, the *spf* promoter (see CRP in Tables 1 and 2), and perhaps the *gal* operon. All of these are CRP-regulated promoters. The *nrd* promoter may also fall into this unusual class, but, as discussed previously, the genetics is not advanced enough to permit a decision. These promoters can be separated into two groups. The

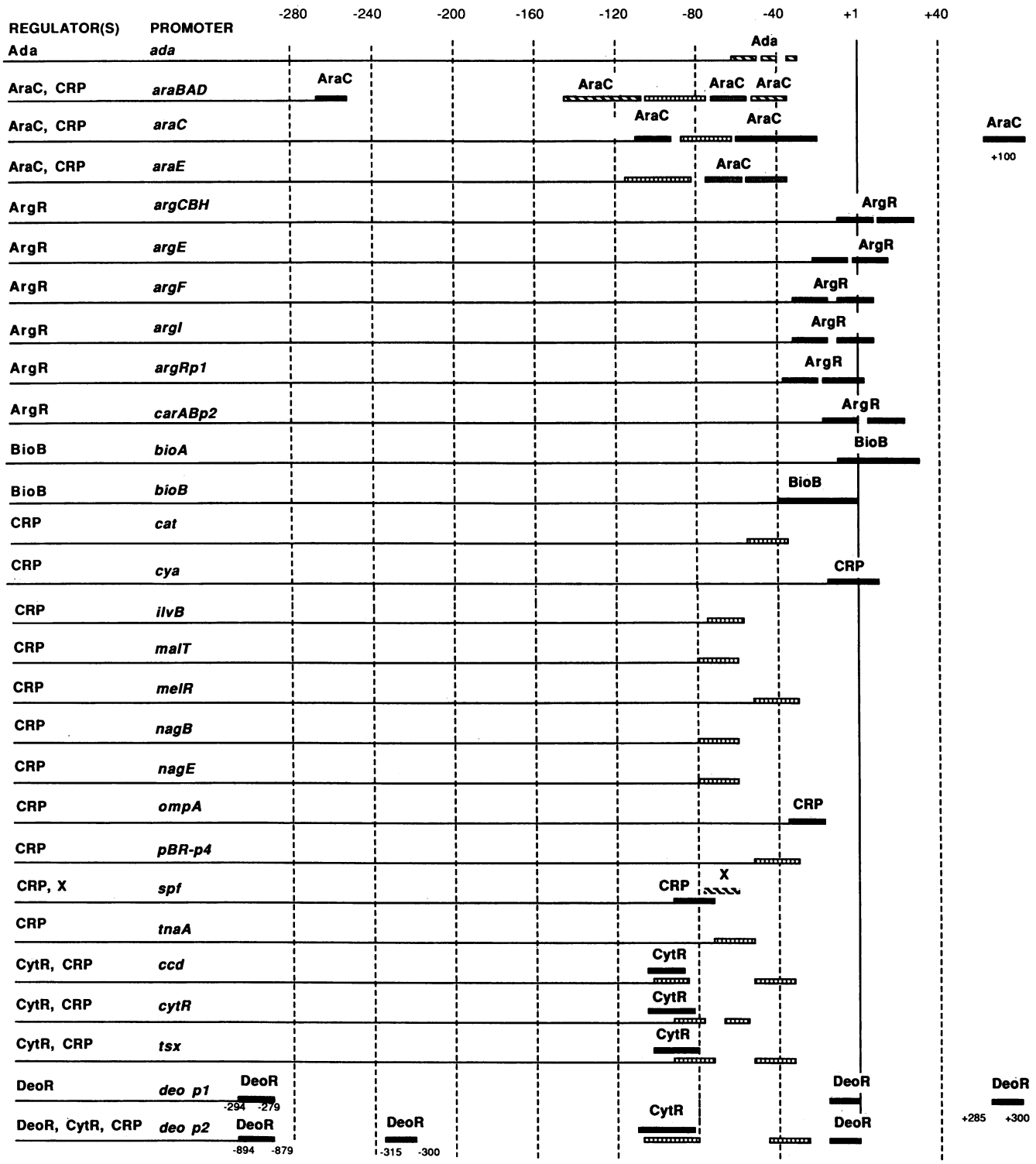

FIG. 1. *E. coli* sigma 70 promoters. Promoters are listed in alphabetical order of regulators. Within regulons they are listed in alphabetical order of promoters, except in cases of multiple promoters which are located one immediately after the other. All the binding sites use the same scale relative to the site of initiation of transcription (+1) marked by a vertical line. Only sites with established regulatory effects on the respective promoter are indicated. Symbols: ▬, known or inferred binding sites of repressors; ▭, binding sites of activators; ▬, protein with a dual positive and negative effect; above each box, an abbreviation that refers to the regulatory protein is indicated, except in the case of positive CRP-binding sites (▭) and positive FNR-binding sites (▭); ▭, domain defined by deletion analysis.

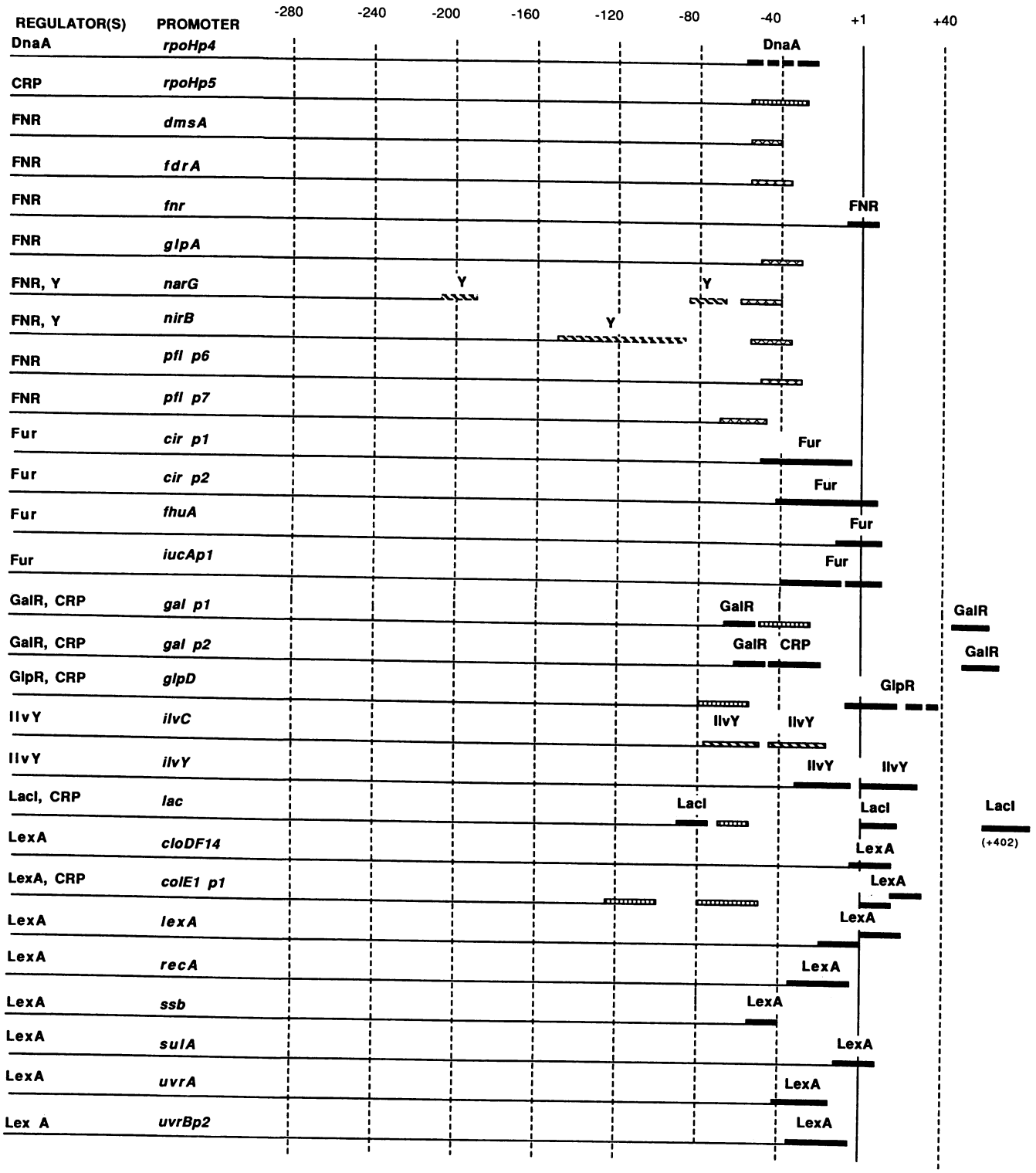

FIG. 1—*Continued.*

CytR regulon and *spf* promoter have a similar arrangement, where the repressor site overlaps a proximal activator site. An illustration of the mechanism of repression comes from studies of the CytR-responsive promoters (143). All of these promoters contain tandem CRP sites, at least one of which is in a proximal position. Upon repression, the CytR repressor binds between the two bound CRP molecules, inactivating them and causing repression. The *gal* mechanism is still controversial, but in one model, in the presence of CRP the repressor binds to an adjacent upstream position to form a

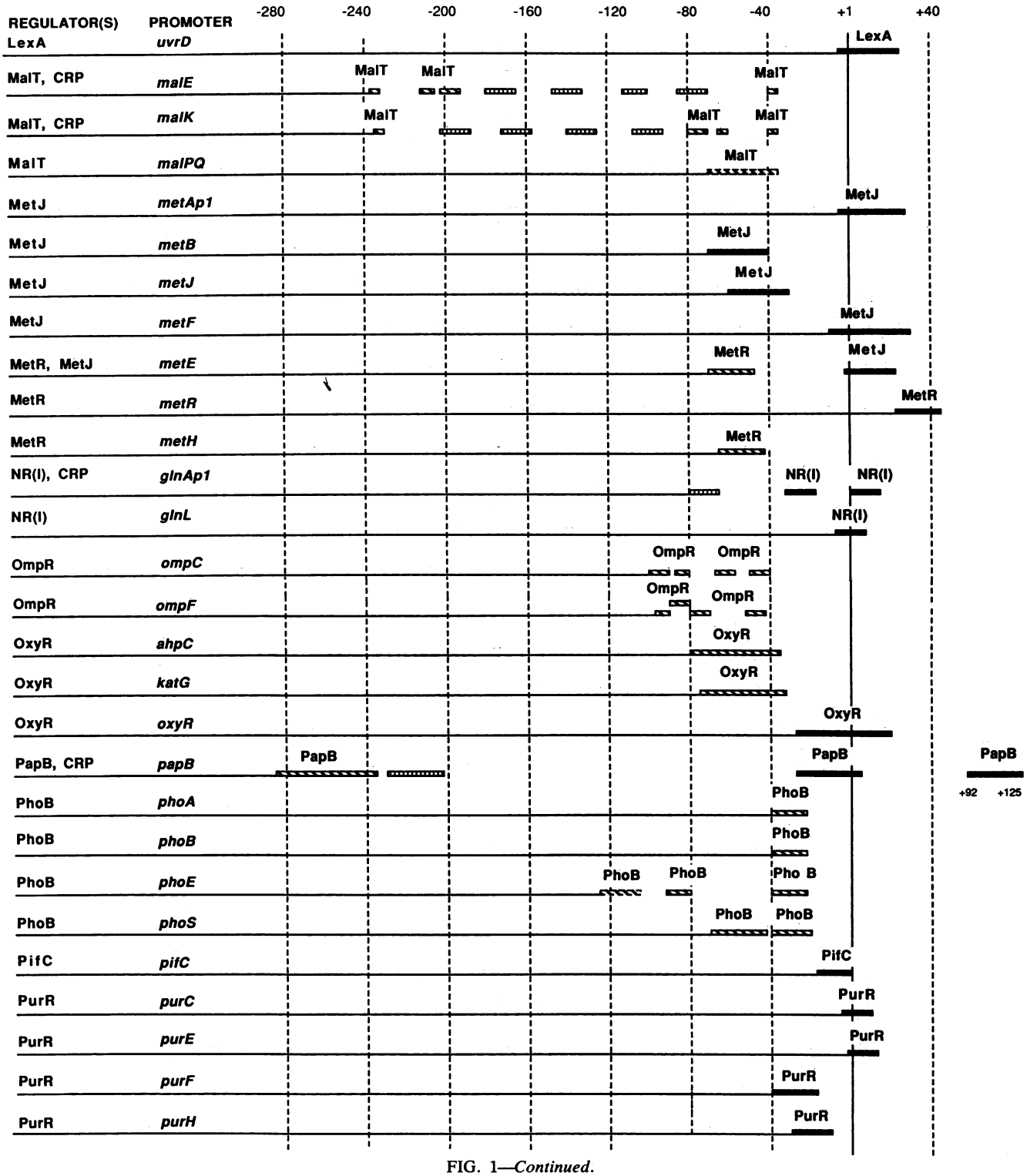

FIG. 1—*Continued.*

repression complex containing both proteins (66). This effectively occludes the promoter, preventing polymerase from acting. In the presence of CRP, transcription proceeds from a site slightly further upstream, which places the operator in a position where it reaches −45 and the polymerase; this repression might then proceed in the conventional manner as discussed above. A remote operator and a repressor capable of multimerizing appear to be required for this repression (85).

Although the positions of operators can vary substantially and there is variation in the way in which repressors work, virtually all repression systems can be explained by repres-

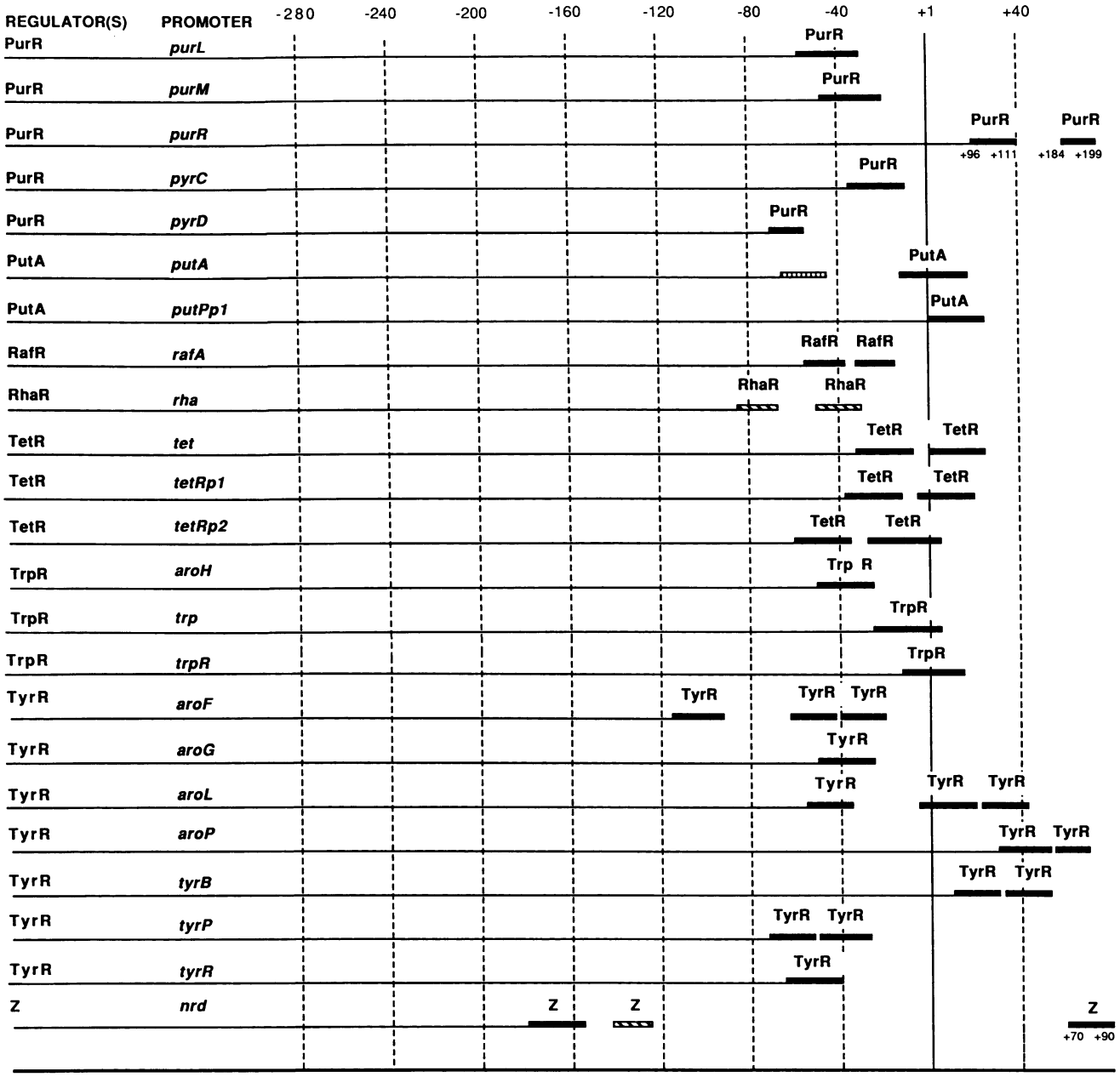

FIG. 1—*Continued.*

sors either touching the polymerase (or its DNA elements) or touching activator complexes, which in turn approach the polymerase.

Duplication of Operators

Among the 76 repressible promoters, 33 (43%) have at least two operator sites that are similar and apparently involved in the binding of the same protein. These promoters are predominantly of two types. Most have duplicated proximal sites that are closely spaced. Eight have the duplicated site as a remote operator well separated from the proximal site. One promoter has site locations that are on the boundary between proximal and remote, and two have only remote sites. There is obviously no general requirement for operators to be duplicated or, when they are duplicated, to be in a particular position. This analysis of the data set leads to two questions: why are operators sometimes duplicated, and why is there no strict requirement for placement of the duplicated operator?

We will first consider the need for duplication. When simple and complex promoters are considered separately, 30% of the former and 56% of the latter are associated with duplicated operators. Thus, there is a tendency for operator duplications to be preferentially associated with complex promoters. This may reflect a need to interact in more than one way with the additional transcription machinery necessarily associated with complex promoters. However, since

TABLE 1. Properties of repressible promoters

Regulator	Promoter[a]	Properties associated with[b]:				Evidence for regulatory sites[c]	References
		Pr	Reg	Op	Act		
AraC	*araBAD*	1	1	R	0	1,2	69, 102, 135
AraC	*araC*	1	1	R	0	1,2	71
AraC	*araE*	0	1	0	0	2	145
ArgR	*argCBH*	1	0	P		1,4	25
ArgR	*argE*	1	0	P		1	25
ArgR	*argF*	0	0	P		4	25
ArgR	*argI*	0	0	P		4	25
ArgR	*argRp1*	1	0	P		2	76
ArgR	*carABp2*	1	0	P		1,2	18, 125
BioB	*bioA*	1	0	0		1	24, 104
BioB	*bioB*	1	0	0		1	24, 104
CRP	*cya*	0	0	0		2	2
CRP	*ompA*	0	0	0		2	97
CRP	*spf*	0	0	0	0	4	109
CytR	*ccd*	0	1	0	P	4	86
CytR	*cytR*	0	1	0	P	1	142
CytR	*tsx*	0	1	0	P	4	86, 142
DeoR	*deop1*	1	0	R		1,2	4, 86, 163, 164
DeoR	*deop2*	1	1	R	P	1,2	4, 86, 163, 164
DnaA	*rpoHp4*	1	0	P		2	167
FNR	*fnr*	0	0	0		4	35
Fur	*cir p1*	1	0	0		2	45
Fur	*cir p2*	1	0	0		2	45
Fur	*fhuA**	0	0	0		4	23, 30
Fur	*iucAp1*	0	0	P		2	16, 29, 30
GalR	*galp1*	1	1	R	0	1,2	1, 85
GalR	*galp2*	1	1	R		1,2	1, 85
GlpR	*glpD*	0	1	P	0	2	174
IlvY	*ilvY*	1	0	P		3	170
LacI	*lac*	0	1	R	0	1,2	36, 101
LexA	*cloDF14*	0	0	0		4	166
LexA	*colE1p1*	1	1	P	P(?)	2,4	33, 34
LexA	*lexA*	0	0	P		2	11, 77
LexA	*recA*	0	0	0		2	11, 77
LexA	*ssb*	0	0	0		4	10
LexA	*sulA*	0	0	0		4	94
LexA	*uvrA*	0	0	0		2,4	127
LexA	*uvrBp2*	1	0	0		2	128
LexA	*uvrD*	0	0	0		2	32
MetJ	*metAp1*	1	0	0		4	92, 160
MetJ	*metB**	1	0	0		1,2	140, 160
MetJ	*metJp1**	1	0	0		1,2	140, 160
MetJ	*metE*	1	1	0	0	1,4	106, 107
MetR	*metR*	1	0	0		1,3	161
MetJ	*metF*	0	0	0		1,4	144
NR(I)	*glnAp1*	1	1	P	0	1,2	119
NR(I)	*glnL*	1	0	0		1,2	158
OxyR	*oxyR*	0	0	0		3	20, 146, 149
PapB	*papB*	1	1	R	R	2	37
PifC	*pifC*	0	0	0		1	64
PurR	*purC*	0	0	0		3	49
PurR	*purE*	0	0	0		3,4	49, 168
PurR	*purF*	0	0	0		1,3	122, 123
PurR	*purH*	0	0	0		3	49
PurR	*purL*	0	0	0		3,4	49, 89
PurR	*purM*	0	0	0		4	89, 141
PurR	*purR*	0	0	R only		1,2	89, 124
PurR	*pyrC*	0	0	0		1,3,4	19, 171
PurR	*pyrD**	0	0	0		4	171
PutA	*putA**	1	1	0	0	2	47
PutA	*putP**	1	0	0		2	47
RafR	*rafA**	0	0	P		3	6
TetR	*tet*	1	0	P		2	52, 53
TetR	*tetRp1*	1	0	P		2	52, 53
TetR	tetRp2	1	0	P		2	52, 53
TrpR	*aroH*	0	0	0		2	65

Continued on following page

TABLE 1—*Continued*

Regulator	Promoter[a]	Properties associated with[b]:				Evidence for regulatory sites[c]	References
		Pr	Reg	Op	Act		
TrpR	*trp*	0	0	0		2	65
TrpR	*trpR*	0	0	0		2	65
TyrR	*aroF*	0	0	P		1,4	22
TyrR	*aroG*	0	0	0		1,2	7
TyrR	*aroL**	0	0	P		4	28
TyrR	*aroP*	0	0	P (?)		1,4	21
TyrR	*tyrB*	0	0	P		1,4	173
TyrR	*tyrP*	0	0	P		1,4	63
TyrR	*tyrR*	0	0	0		4	173
Z	*nrd*	0	1	R	R	1,4	156, 157

[a] Promoters are listed in alphabetical order of repressor. For promoters followed by an asterisk, the initiation of transcription is assigned by reference to the −10 region.

[b] Abbreviations: Pr, properties associated with promoter, indicating whether a promoter is either multiple (shown by 1) or single (shown by 0); Reg, properties associated with regulation, indicating whether a promoter is subject to multiple (shown by 1) or to only one (shown by 0) system of regulation; Op and Act, properties associated with operator and activator, indicating whether there is a single site (shown by 0), a proximal duplication (shown by P), a remote duplication (shown by R), or no site (blank). A (?) symbol indicates when the properties assigned are ambiguous.

[c] The evidence that supports the existence of regulatory sites is classified as follows: 1, mutational evidence; 2, specific binding of purified protein; 3, specific binding of partially purified protein or using cellular extracts; 4, similarity with consensus sequence. For promoters subject to multiple regulation, evidence concerning negative sites is indicated here and evidence indicating positive regulation is indicated in Table 2. Information on the following promoters were obtained with *Salmonella typhimurium*: *carAB*, *metB*, *metJ*, *metE*, *metR*, *putA*, and *putP*. In the few cases when it is known that the promoter has a similar array both in *E. coli* and in *S. typhimurium*, it appears only once in the table. Potential repression of *ompC* by IHF is not included (58).

many complex promoters have a single operator and since many simple promoters involve duplications, there seems to be no obligatory regulatory theme involving operator duplications.

If there is no general theme or requirement, why are operators duplicated in about half of the promoters? There are a few well-characterized cases in which the function of duplication is known. The best of these are the *lac* and λ operators. For *lac*, the function of the extra operators is primarily to strengthen the binding of the repressor (36, 43, 101, 130). This is achieved through cooperative interactions between separated binding sites. A part of the function of the duplication in λ is also to strengthen binding by the repressor by cooperativity (111). Thus, at least one rationale for operator duplication is to allow for stronger repressor binding through cooperative interactions.

A second rationale is also suggested by these systems: operator duplications allow for a flexible response to changing physiological conditions. Thus, for *lac*, the extra operators come into play mostly when cellular conditions require the most severe repression (36). For λ, the duplications allow the level of the regulated protein to be maintained within narrow limits even under widely varying physiological conditions (111). Thus, operator duplications can provide both a tighter regulation and a more flexible response of the regulatory apparatus. For this to be effective, however, the analysis indicates that at least one of the operators must be in a proximal position, where it is presumed to touch the polymerase and interfere with its function.

If operator duplication can aid function, is the location of the duplicated site important? Of 33 promoters with duplicated operators, 23 are proximal duplications (including the marginal *aroP*) and 10 are remote duplications (including the marginal *nrd* and *purR*); i.e., two-thirds of the extra operators are proximal and almost one-third are remote. Several studies have shown that as operators are moved apart, the cooperative interactions between them diminish (initially shown in reference 56). Thus, the location of two-thirds of the duplicated operators in a nearby proximal position represents an arrangement in which the strongest assistance can be given to repressor binding. One expects that the enhanced stability achieved in these cases is important for maximizing the level of repression. There may be additional assistance from increasing the potential for contact between the cooperatively bound repressors and RNA polymerase.

This view is challenged somewhat by the analysis showing that duplications occur more frequently in complex promoters than in simple promoters (Table 4). Complex promoters need not be more tightly repressed but must respond to multiple signals, and each of the operators could in principle play a unique regulatory role. In that case, the role of duplication would be principally increased flexibility rather than increased repression. It may be that duplications in multiple promoters provide more flexibility whereas duplications in simple promoters allow for tighter repression as well as flexibility.

The dominant role of flexibility receives some support from studies suggesting that promoters within a regulon controlled by single operators can be at least as tightly repressed as other promoters containing duplicated operators. The comparative evidence on this issue is still sparse. For example, within the LexA regulon a promoter containing a duplication requires more repressor for tight repression than one with a single operator (11, 77). In TyrR-controlled promoters there is an unusually wide variation in sequences of operators (173), implying that individual sites may be of relatively low affinity. In these cases the duplicated operators may be of somewhat weaker individual affinity and compensate for this by duplication. Thus, a comparable potential for tight repression is achieved while also building in the flexibility to use the two sites for independent regulatory events, such as occurs for lambda Pr. We will not know whether the need for duplication is dominated by its association with flexibility rather than high affinity until more quantitative information on repression is available.

The data collected in Table 1 indicate another determinant of whether operator duplications will occur, i.e., the nature of the repressor protein involved. Thus, all the ArgR-regulated promoters have proximal duplicated operators irrespective of the organization of promoters; none of the

TABLE 2. Properties of activatable promoters

Regulator[a]	Promoter[a]	Properties associated with[b]:				Evidence for regulatory sites	References
		Pr	Reg	Act	Op		
Ada	*ada*	0	0	0		1,2	126
AraC	*araBAD*	1	1	0	R	1,2	59, 70
AraC	*araC*	1	1	0	R	1,2	48
AraC	*araE*	0	1	0	0	2	145
CRP	*cat*	0	0	0		2	72
CRP	*ccd*	0	1	P	0	2,4	165
CRP	*colE1p1*	1	1	P(?)	P	1,2	139
CRP	*cytR*	0	1	P	0	1,2	142
CRP	*deop2*	1	1	P	R	1,2	142, 162
CRP	*galp1*	1	1	0	R	1,2	60
CRP	*glpD*	0	1	0	P	2,4	174
CRP	*glnAp1*	1	1	0	P	4	119
CRP	*ilvB*	0	0	0		4	39
CRP	*lac*	0	1	0	R	1,2	87
CRP	*malT*	0	0	0		1,2	88
CRP	*melR*	0	0	0		1,2	169
CRP	*nagB*	1	0	0		4	108
CRP	*nagE*	1	0	0		4	108
CRP	*pBRp4*	0	0	0		2,4	113
CRP	*putA*	1	1	0	0	4	47
CRP	*rpoHp5*	1	0	0		4	99
CRP	*tnaA*	0	0	0		2,4	27
CRP	*tsx*	0	1	P	0	4	142, 165
FNR	*dmsA*	0	0	0		4	35
FNR	*fdrA*	0	0	0		1,4	35, 61
FNR, Y	*glpA*	0	0	0		4	35
FNR, Y	*narG*	0	0	0		1,4	35, 75
FNR	*nirB*	0	1	0		1,4	35, 61, 62
FNR	*PFLp6*	1	0	0		4	133
FNR	*PFLp7*	1	0	0		4	75, 133
IlvY	*ilvC*	1	0	P		3	170
MalT	*malE*	1	1	R		1,2	114, 115
MalT	*malK*	1	1	R		1,2	114, 115
MalT	*malPQ*	0	0	0		1,4	138
MetR	*metE*	1	1	0	0	1,3	161
MetR	*metH*	0	0	0		4	160, 161
OmpR	*ompC*	0	0	P		1,2	80, 81, 154
OmpR	*ompF*	0	0	P		1,2	38, 116, 154
OxyR	*ahpC*	0	0	0		3	146
OxyR	*katG*	0	0	0		3	146, 149
PapB	*papB*	1	1	R	R	1,2	37, 41
PhoB	*phoA*	0	0	0		4	84
PhoB	*phoB*	0	0	0		1,4	84, 154
PhoB	*phoE*	0	0	P(?)		1,4	153
PhoB	*phoS*	0	0	P		2,4	83
RhaR	*rha*	0	0	P		2	150–152
X	*spf*	0	1	0	0	2,4	109
Z	*nrd*	0	1	R	R	1	156, 157

[a] Promoters are listed in alphabetical order of activator. *araBAD*, *araC*, *araE*, and *papB* promoters are also activated by CRP; *narG* is also activated by nitrate. Potential activation of *pifC* by IHF is not included (64). The unknown activator of *spf* is named X. The nitrate-dependent activator of *narG* and *nirB*, probably NarL, is named Y, and the unknown regulator of *nrd* is named Z. In all the cases, references refer only to activator sites.

[b] See Table 1 footnotes for an explanation of abbreviations and symbols.

MetJ-regulated promoters have duplications irrespective of being singly or multiply organized and multiply regulated. Similarly, all the promoters controlled by the TyrR repressor, as well as those controlled by TetR, have proximal duplications, whereas most of the many promoters under LexA repression, as well as all the promoters of the *pur* regulon except one, have a single operator; the *lexA*, *colE1* (see LexA in Tables 1 and 2), and *purR* promoters constitute the few exceptions. It appears that the ability to function primarily in association with either single or multiple promoters is built into the structure of certain repressors. One might expect this to be related to the ability of the protein to multimerize so that the duplicated sites may be bound cooperatively, but the generality of this idea has not been tested.

Thus, the analysis indicates that there are two situations in which there is a stronger potential need for operator duplication. One is when the promoter is complex, and the other is when the particular repressor involved is not designed to function effectively by using a single site. These needs can be rationalized in terms of needing greater flexibility and tighter repression.

TABLE 3. Regulation in multiple promoters

Promoters[a]	Orientation of promoters[b]	Duplication of regulatory sites[c]	No. of systems[d]	Observations
araBAD,C	D	R	2	Shared sites
argCBH,E	D	P	1	Shared sites
argR p1,p2	P	P	1	One Pr insensitive
carAB p1,p2	P	P	1	One Pr insensitive
bioB,A	D	0	1	Shared site
nagE,B	D	0	1	Shared site
deo p1,p2	P	R	3	Shared sites and different systems on each Pr
rpoH p5,p4	P	0	2	Different systems on each Pr
PFL p6,p7	P	0	1	Independent sites on each Pr
cir p1,p2	P	0	1	Shared sites
gal p1,p2	P	R	2	Shared sites and different effects on each Pr
ilvY, ilvC	D	P	1	Shared sites and different effects
colE1p1	P	P	2	One Pr insensitive
uvrB	P	0	1	One Pr insensitive
malE,K	D	R	2	Shared sites
metA p2,p1	P	0	1	One Pr insensitive
metB,J	D	0	1	Shared sites
metE,R	D	0	2	Shared sites
glnAp1	P	P	2	Different sigma factors
glnL,A	P	0	2	Pr internal to an operon
papB,I	D	R	2	Regulation on only one Pr described
putP, putA	C	0	2	Shared sites and different system on one Pr
tet, tetR p1,p2	D	P	1	Shared sites

[a] The promoters are ordered as they appear in Fig. 1.
[b] Abbreviations: P, parallel; D, divergent; C, convergent.
[c] Abbreviations: P, proximal; R, remote; 0, absent.
[d] The total number of systems of regulation that regulate the set of multiple promoters is indicated for each group. On the basis of RNA polymerase-binding sites, *bio* and *tet* could be considered convergent promoters.

Remote Operators

Table 4 shows that one category of promoters stands out as quite different from the others with respect to operator organization. Of the 10 promoters that are both multiple and subject to multiple systems of regulations, 8 have duplicated operators and 6 of these are in remote positions. At the other extreme are the simple promoters, in which a remote operator duplication occurs only once (*purR*) among the 37 examples (and this may not work at initiation [see above]) and proximal duplications occur in fewer than one-third of the cases. One might argue that the proximal regions around the very complex promoters will by necessity be crowded with regulatory signals, accounting for the high occurrence of remote operators. Proximal duplication requires building a regulatory region with two sets of proximal RNA polymerase elements, at least two negative elements to affect both promoters and at least one site for positive regulation, such as a CRP site. Since, as discussed below, positive sites generally must occur in the proximal region, there may be little room to build in duplicated proximal operators. One of the operators apparently must be placed within the proximal region (see above), but there may simply be no room to build in a second one without destroying the various other recognition elements.

Despite these restrictions, the analysis shows that 74% of proximal duplicated operators are located downstream of −40, in principle leaving enough room for a CRP site located as in *lac*, and 9 of these 23 cases could have a CRP site similar to the one of *gal*. Thus, it is possible to build promoters with two proximal operators and a CRP site, but there are only two examples in multiple promoters subject to multiple regulation, *colE1* and *glnAp1* (see LexA and NR(I) in Tables 1 and 2). Recall that duplication preferentially occurs in a remote position in complex promoters subject to multiple systems of regulation (Table 4). The requirements to build in another set of polymerase recognition elements in these cases may make proximal operator duplications rare. Some of this may be understandable in evolutionary terms, as discussed below.

The need for duplication in these cases may be understood in terms of both flexibility and affinity. Obviously, in the most complicated promoters the need for flexibility is great. Moreover, given the crowded regulatory region, it may be difficult to build in even one proximal operator with a sequence compatible with all the other requirements. This difficulty can be overcome, as evidenced, for example, by the operator in MetE, which is not duplicated and overlaps the −35 region of a divergent promoter. However, such overlapping recognition may generally require that the proximal operator be of low affinity and demand the additional assistance from a duplicated operator.

It is useful to consider the individual cases in which remote duplications occur. There are 10 promoters with operators duplicated in remote regions; these represent only 13% of the repressible promoters. Two of these are the marginally remote *aroP* and *nrd*, cited above, for which the regulatory mechanism is not fully established in the sense that there may be connections to other proximal elements. Another unusual case is the *purR* negative autorepression that may be regulated at the level of elongation. The remaining eight cases are the *lac*, *deop1*, *deop2*, *galp1*, *galp2*, *araC*, *araBAD*, and *papB* promoters.

As just discussed, these promoters with remote duplications are mostly complex promoters subject to multiple systems of regulation. The prominent exception is the *lac* promoter, which has a well-known overlapping promoter

TABLE 4. Duplication of operators in simple and complex promoters

Type of promoter[a]	Operator duplication[b]			Total
	R	P	None	
Complex, multiple promoters	1	9	11[c]	21
Complex, multiple regulation	2	2	4	8
Complex, multiple promoters and regulation	6	2	2	10
Simple	1	10	26	37
Total	10	23	43	76

[a] Definitions of the type of promoters are given in the text.
[b] Abbreviations: R, remote; P, proximal. Data are taken from Table 1. *aroP* is counted as proximal duplication, *nrd* and *purR* are the only remote duplications in promoters with no proximal sites as mentioned above.
[c] *metR* is included here.

that is active in vitro but apparently not in vivo. This concentration of remote operators within multiple promoters compares to only 37% of the total repressible promoters being defined as multiple (recall that multiple promoters are those that share a regulatory region or transcribed gene with another promoter).

The second striking grouping is that seven of the eight promoters with remote duplications of operators (*deop1* being the exception) contain nearby interaction sites for CRP. This compares to 19% of repressible promoters containing CRP sites involved in activation. Taken together with the above comparison, it is clear that remote operator duplications are concentrated in multiple promoter regions that are coregulated by CRP. We conclude that remote operator duplications are usually associated with multiple promoters subject to more than one system of regulation. This conclusion is based on only eight examples and, of course, could possibly be subject to modification. The reason for this is not only the relatively small number of examples but also the possibility that, as investigation of existing examples intensifies, previously hidden remote operators will be identified. The only examples of multiple promoter regions coregulated by CRP that do not have remote operators are *glnAp1* and *putA*. The *glnAp1* promoter is not a true exception since the other promoter is actually transcribed under the control of another sigma factor, sigma 54, leaving the *putA* case as the only apparent current exception to the rule that multiple promoter regions coregulated by CRP must have operator duplications (see Addendum in Proof).

One expects that such complex promoters evolved from more simple systems, and it is informative to consider how this may have led to the current organization of complex promoters. For the *lac* promoter, it has been argued that the original promoter was the now vestigial upstream promoter p2 (87). This promoter is not catabolite controlled and overlaps the upstream *lac* O_3 operator in a position that suggests that it was negatively regulated by the *lac* repressor bound to this operator. If evolution of the catabolite control apparatus occurred after origination of the *lac* promoter, as seems likely, the current CRP-dependent promoter would have been built in a downstream position overlapping the original promoter. Eventually, the original promoter may have lost function; the upstream operator is also now primarily vestigial (101). This may be seen as somewhat similar to the *gal* case, in which at the present stage of evolution both the upstream and downstream promoters retain function. In both cases there are remote downstream operator duplications which may have been necessary to retain tight repression as the catabolite-dependent proximal promoter-operator sequences evolved.

Evolutionary arguments aside, the crowding of elements within these multiple promoters is evident in the various cases of remote operator duplications. This argument has been made explicitly for the *araBAD* system, in which a compact regulatory region controls the divergent *araBAD* and *araC* promoters (134). Both *araBAD* and *araC* have proximal sites for the binding of AraC protein, and a remote site exists as well. There are also built-in sites for activation by CRP and the AraC protein in its role of activator. The complex regulation of this system involves differential cooperative protein interactions among the different sites. It is difficult to imagine the evolution of such a complex and compact system without the participation of remote regulatory sites.

One might ask why operators do not appear in remote positions more frequently, since they obviously can function at a distance. Most probably the answer is that the remote placement has certain disadvantages. If the two sites are to cooperate, the repressors bound to them must be capable of binding each other, looping out the intervening DNA. Essentially this means that the repressors, which normally contain a single type of polypeptide, must be constructed so as to be capable of multimerizing. Although this would also have to be true for proximal duplications, it has been established that cooperative interactions usually diminish with distance, making the remote placement less effective. Thus, the preferred solution would be to place the operators near each other to maximize the efficiency of repression, but since this may sometimes require a very precise placement in a crowded region, remote operator placement occurs in some cases.

Lastly, one might ask why it is particularly CRP, and not positive regulators in general, that most often is associated with the remote operator duplication. The answer to this may come from a rather unexpected property of the data base: the large majority of repressible promoters that are also subject to activation have CRP as the activator (Fig. 1). Thus, there may be nothing special about the involvement of CRP; it may simply be that physiology has evolved to demand activation in addition to repression principally for catabolite-sensitive operons.

POSITIVE REGULATION OF SIGMA 70 PROMOTERS

Proximal-Site Position

Of the 107 sigma 70 promoters, 48 are activatable. Of these, 47 have proximal sites, meaning that there is only one potential exception to the general rule that regulation must proceed via at least one proximal element. The sole exception, *spf* (see CRP in Tables 1 and 2), is in fact just marginally remote, approaching the −70 position. Recall that the definition of a proximal site (−65 to +20) is one that overlaps with the binding site for RNA polymerase or is in the same position as the CRP site as defined in the *lac* operon. This definition was used because there is considerable evidence that CRP can touch the polymerase in this arrangement (see above). This analysis indicates that activation, like repression, generally proceeds through a mechanism involving at least one proximal site. The rationale for this is that the polymerase is normally touched by the activator when its properties are to be altered.

One difference between the distribution of proximal sites

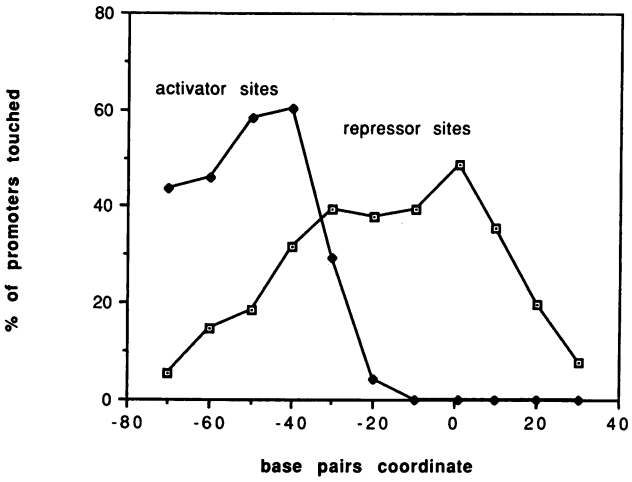

FIG. 2. Plot demonstrating the fraction of promoters with regulatory sites that touch the indicated promoter positions. The data was obtained by noting whether each of the regulatory sites from Fig. 1 overlapped each of the promoter positions. Thus, for example, 60% of activatable promoters have sites overlapping −40, and 49% of repressible promoters have operators overlapping the start-site at +1. AraC sites in *araE* were excluded.

for activators compared with repressors is that the proximal activator sites are located in a much less variable position. This is emphasized by the observation that 30 of the 47 proximal sites touch the −40 position. Almost all of the exceptions involve CRP. Remarkably, this relative uniformity in position is seen in a collection of promoters that represent a sample of 13 different activator proteins. Figure 2 shows a comparison of the distribution of positions touched by activator sites compared with repressor sites within the proximal region. Note that the activator sites are restricted to a smaller region than the repressor sites. In general, proximal activation sites are located to overlap the promoter −35 region, or, in the case of CRP, alternative upstream in-phase positions are sometimes used.

When regulons are considered as a class, the fixed position of proximal activation sites compared with repression sites is even more striking. For example, activation sites for fumarate and nitrate regulator protein FNR have been identified in seven unlinked promoters; six are either at −50 to −40 and one is near −60. Similarly, the four promoters controlled by PhoB have their most proximal sites in an identical position centered near −30. This is in marked contrast to the negative control of regulons, discussed above, where the operators occupy positions throughout the acceptable part of the proximal region, as defined by Fig. 2.

There are many examples of proximal CRP sites, and these are slightly more variable but still much less so than operators; the CRP sites fall primarily into three categories at the *lac*, *gal*, and *malT* positions. Of the 25 CRP sites, 8 fall around −40 on the *gal* position, 6 are around −60 as in *lac*, and 6 are centered at −70 as in the *malT* promoter. The remaining sites in a different position are those of *malE*, *malK*, *araBAD*, *araE*, and *papB* promoters. In these promoters, as has been shown for the *mal* promoters (115), CRP may activate in concert with additional proteins bound to proximal positions.

A clue to the more fixed positions of activation sites comes from studies in which the activator sites have been artificially moved to other positions. Thus, when CRP sites are moved to different proximal positions, the activation is strongest from the *gal* position, still strong from the *lac* position, and detectable from the *malT* position. CRP was not able to activate substantially from other positions (40). This corresponds to the natural distribution of sites in the data base, in which CRP sites are restricted primarily to near these three positions.

The three acceptable CRP positions are separated by an integral number of DNA helical turns, suggesting that CRP must be positioned stereospecifically with respect to the bound RNA polymerase. From two of these positions the CRP-RNA polymerase contact could be similar since the upstream half of the CRP dimer in one position overlaps the downstream half of the CRP dimer in the other position. In the furthest position, centered near −70, a different means of touching would have to be used. This presumably involves the same site on RNA polymerase since it is restricted to the same helix face. Recall that, with the exception of these few −70 CRP sites, most activatable promoters have sites that touch −40, as does the strongest CRP site. This suggests that the large majority of activator sites are placed so that they touch the polymerase in such a way that high activation is achieved.

The most obvious exception to this is the existence of the several CRP sites in the *malT* position. Studies of the activation of *malT* by CRP and of *ompF* by OmpR give clues to how more distant activators can regulate RNA polymerase (81, 88). The positions of the OmpR sites have been altered systematically, and the effects on activation have been measured. OmpR works best from its native proximal position and less well as the sites are moved upstream. As in the CRP studies, the activation is maximal when the OmpR sites are restricted to the same side of the helix as the natural site. These results yet again suggest the need for a stereospecific contact between activator and RNA polymerase. Since there is residual activation from upstream, it is possible that the intervening DNA is bent to allow the contact from upstream; the energy cost in bending the DNA to bring the proteins together would account for the lower activation from the upstream positions. Similar reasoning may account in part for the lower CRP activation from the upstream *malT* position. In addition, it has been suggested that CRP activation at *malT* occurs by a unique mechanism. Instead of assisting in promoter recognition, CRP appears to promote escape and chain initiation by the already bound polymerase (88).

Thus, the few activator sites that do not touch near −40 may nevertheless involve a similar contact with RNA polymerase. One may ask why this particular position near −40 is so common. The mechanism of promoter recognition provides a plausible rationale (44). In the absence of activators, the process of promoter recognition involves polymerase binding, presumably simultaneously, to the recognition elements at −35 and −10. Studies involving activation of regulons have shown that the −35 elements are often far from consensus, indicating weak intrinsic promoter-binding activity (26). In a sense, the role of the activator in these cases may be to replace the −35 element and provide a substitute contact point for RNA polymerase, as has been proposed for the *pho* regulon (83).

Note that it is almost always the −35 region that is approached by the activator. The analysis shows that activation sites never coincide with the −10 promoter element (Fig. 2). The reason for this is probably that the −10 region has an additional critical role in promoter activation. DNA

TABLE 5. Regulatory proteins with repressor and activator effects

Protein[a]	Secondary function
AraC	Dual effects on *ara* regulon
CRP	Represses *spf*, *cya*, *ompA*, *galp2*
FNR	Represses *FNR*
IlvY	Represses *ilvY*
MetR	Represses *metR*
NR(I)	Represses sigma 70 promoters and activates sigma 54 promoters
PapB	Dual autoregulatory effects

[a] OmpR is excluded because there is no known natural promoter repressed by it.

melting, to expose the bases that must be read during transcription initiation, begins within this region. If the activator, having recognized the double-stranded DNA structure, were to be bound here, it would be necessary to dissociate the activator to accomplish the essential strand opening. In fact, the data base analysis shows that no activator binds anywhere within the region that must be melted (Fig. 2). Thus, if the activators are to substitute for the lack of contacts during promoter recognition, but not interfere with the critical strand opening, by far the preferred arrangement is the one observed, i.e., binding near where polymerase normally touches down at −35. The closest activator sites to the −10 region of sigma 70 promoters are the *c*II activator, which reaches position −15 of the *p(I)* promoter in phage lambda (55), and the MerR activator, which binds from −36 to −10 in the *merT* promoter of a transposon isolated in *Pseudomonas aeruginosa* (78, 103) (these are not included in Fig. 1, because the catalog is restricted to *E. coli* promoters).

Activators That Can Also Repress

Certain of these activator proteins can also mediate repression of a small number of promoters. The examples of this type from the data base are collected in Table 5. These are CRP, repressing several promoters, and the activators FNR, MetR, PapB, and IlvY, repressing their own synthesis. The first and most obvious classification here is that with the exception of CRP, they are all repressing their own synthesis. The special case of NR(I) (NtrC) is also included in the table as it represses its own synthesis from a sigma 70 promoter but activates it from an overlapping sigma 54 promoter. The advantage of such autoregulation is to keep production of the regulator within narrow limits, as has been reviewed previously (111). The observation that many proteins can act as either repressors or activators emphasizes that the position of DNA binding may dominate the individual properties of the protein in determining whether activation or repression occurs.

Analysis of the data base indicates that in some cases one might predict whether a potentially bifunctional protein will activate or repress from knowledge of the position of the binding site. The clearest examples of this are the regulons that have the largest numbers of examples, those controlled by FNR and CRP. Recall that the site from which FNR activates the seven known promoters of its regulon is fixed near −40 or −50. In the single example of FNR acting as a repressor, the binding site is located near +1. As discussed above, this position is out of the region where activation generally can occur (Fig. 2). Thus, in this case FNR binding leads to repression, probably via the usual mechanism of touching polymerase or occluding its binding site. To restate this idea, a change in binding site location can convert an activator into a repressor.

A similar argument can be made for CRP, but the predictive value is not as great. Recall that CRP activation sites are mostly restricted to three phased position. When CRP represses at the *ompA* and *cya* promoters, it is bound outside of its activation position and within the common zone of repression (Fig. 2). Neither repression site touches the critical −40 position, *ompA* being repressed from the out-of-phase position near −30 and *cya* from within the exclusive zone of repression overlapping the −10 position. These would constitute easily predictable cases.

In the two remaining cases of CRP repression, it is less easy to predict, from knowledge of position alone, whether CRP represses. This may be because these cases are special in that the CRP sites overlap with those of another regulatory molecule and appear to repress by influencing the properties of that molecule. In the *gal p2* promoter the CRP site does not activate, probably because it is 5 bp out of phase with the three common locations from which it is capable of activating (40). This position, however, is within the general zone of both activation and repression (Fig. 2). CRP may repress in this case by cooperating with the *gal* repressor which binds immediately adjacent to form a repression complex (66). At the *spf* promoter, the CRP site near −80 can assist in repression apparently because it overlaps with the site at which the gene-specific activator works.

These examples indicate that activators can act as repressors when they are outside of the normal activation locations, but can repress under other influences as well. Thus, AraC and IlvY have in common that their function is influenced by the binding of small molecules to the proteins. The biochemistry of these interactions is still a bit controversial, but they may be explained by induced conformational changes that form or disorganize protein domains required for activation (90). Thus, it may be that activators must both be located in the appropriate position and have their structured activation domains exposed in order to activate. Otherwise repression might occur.

A final consideration here is that even if these criteria are met, the way the regulator functions may depend on how the promoter itself is constructed. For example, if activators work by helping the binding of RNA polymerase to DNA, then positioning an activation site near a promoter that already directs strong binding may not be effective. In an extreme example of this type, the OmpR regulator has been shown to repress and activate from the very same position (155). It activates intrinsically weak promoters by promoting RNA polymerase association, but when the promoter elements are improved to allow stronger intrinsic binding, it actually represses. Apparently the same way of touching the polymerase that holds it near the weak promoter also holds it near the strong promoter but, in the latter case, slows it from proceeding further in the transcription initiation pathway.

Overall, analysis of the data base indicates that these few cases are interesting exceptions and not the rule. By and large, natural promoters have their activators positioned appropriately for activation, generally near the promoter −35 region. Activation often depends on physiological induction of the protein itself or on small regulators to induce the appropriate protein conformation. When it is desirable for activator proteins to be used for repression, generally in

autoregulation, the sites often appear in positions appropriate for repression rather than activation.

Promoters with More than One Activation Site

Among the 48 promoters subject to activation, 17 involve more than a single activation element. (The *ada* and *rha* sites are not considered duplications but unique discontinuous elements.) Of these 17, 13 are duplications of the same activation site. Eight of these duplicated sites are in the proximal position: *ccd*, *cytR*, *tsx* (see CytR in Tables 1 and 2), *deo p2*, *ilvC*, *ompC*, *ompF*, and *phoS*. Four of the homologous duplications occur in remote positions: *malE*, *malK*, *papB*, and *phoE*. Some cases involve more than one activator. We will consider the homologous duplications first.

Four of these eight cases are duplications of the CRP site in promoters of the CytR regulon (143). The position of these sites is rather fixed, with the proximal site centered in the usual position near −40 and the adjacent site centered near −90. These duplications appear to be related to a highly specialized aspect of CytR regulation. The CytR repression site extends from −30 to −100, but CytR repressor cannot bind this DNA on its own. Only when CRP occupies the two sites can CytR bind and repress. Thus, in a sense, the tandem CRP sites can be formally considered part of the repression apparatus. There is no evidence as yet for the involvement of the upstream site in activation.

The remaining homologous duplications are of the PhoB sites at *phoE* and *phoS*, the IlvY sites at *ilvC*, the OmpR sites at *ompC* and *ompF*, and the CRP sites at *colE1* (see LexA in Tables 1 and 2). In all of these cases the repeated sites are arranged in a manner that would in principle allow the bound proteins to interact. In the *phoS*, *ilvC*, and *omp* cases the sites are directly adjacent to each other. Adjacent sites for the lambda *c*I protein acting as an activator bind the protein cooperatively; such cooperative binding has several regulatory advantages in principle (111). In the *colE1* and *phoE* cases the tandem sites are separated by 20 and by 40 bp, respectively, implying that they occur on the same face of the helix. Yet again, in comparison with the lambda example, it has been shown that cooperative binding can occur when the sites are moved apart, but only if they stay on the same helix face (56). Although definitive studies have not yet been done on these systems, it is possible that the duplications allow cooperativity and hence a stronger response. It is also possible, but unprecedented, that the duplications enhance activation simply by providing more than one site from which activation can occur.

The remaining cases with more than one element are complex because they contain sites that bind more than one protein. The examples of promoters of this type are principally a small subset of those controlled by either the AraC protein or the MalT protein and possibly NarL (indicated as Y in Fig. 1) involved in nitrate-nitrite metabolism. We begin with the *malE* promoter because it has been characterized in detail.

The *malE* case has already been discussed at a simpler level in that CRP can activate from a position upstream from the normal proximal zone. In addition, the *malE* promoter contains a site for MalT activation right at −40. This case has been studied extensively, and it has been shown that the upstream CRP site cooperates with a series of a total of seven MalT or CRP regions to allow formation of a large multiprotein activation complex (114, 115). All of these sites except one are phased so that the proteins bind the same DNA helix face. This results in a wrapped nucleoprotein complex that approaches the polymerase closely at the most proximal site near −40. Thus, this differs from the simple cases discussed above only in the more complex means of delivering the activator to near the −40 region of the promoter.

A similar argument can be made for the two promoters controlled by upstream CRP sites and proximal AraC sites. In the *araBAD* and *araE* promoters, the position of the CRP site is the same, centered near −100. According to the above arguments, it is unlikely that CRP bound there could easily touch the polymerase directly. In both promoters the AraC protein-binding site is located in an immediately adjacent proximal position, which in fact overlaps the prominent −40 position. These cases would be brought into harmony with the theme of touching polymerase near −40 by presuming that CRP cooperates with AraC protein to form a DNA-bound complex that touches polymerase near the usual position.

There are three remaining cases in which heterologous activation sites reside in a remote position. Intriguingly, two of these cases involve nitrite control in cooperation with proximal FNR sites. Since these FNR sites are in the usual −40 position, it is easy to imagine cooperation between the nitrite regulator and FNR at −40, near to the appropriate polymerase position. The unusual aspect here would be that it would require cooperation between different proteins, presumably by looping over a considerable distance. Looping involving identical proteins occur with repressors, but the heterologous mechanism has been thought to be associated primarily with the sigma 54 transcription apparatus in *E. coli*. Unfortunately, the DNA regions between these remote sites and the proximal FNR regions have not been characterized fully genetically, and the putative activator has not been characterized either genetically or biochemically (35, 61, 75), so one cannot determine whether these cases truly differ from the *malE* or other models. The final, somewhat unusual case is the *papB* promoter, in which adjacent CRP and PapB sites activate from a far-remote position. In this case there is another PapB site in a proximal position centered near −15, so mechanisms involving delivery of the remote upstream sites to the region near RNA polymerase can still be postulated.

Although these few cases are intriguing, this discussion should not detract from the main point revealed by the analysis. This is that virtually all cases of activation have an arrangement whereby the polymerase can plausibly be touched by the activator. The means of delivery of the activator to the polymerase may, of course, differ among promoters. This provides a pleasing harmony in regulation, since, as discussed above, a similar mechanism appears to apply to regulation via repressors. Whether regulation is positive or negative, the arrangement of regulatory sites seems to be designed for ease of communication with the proximal binding site for RNA polymerase.

SIGMA 54 PROMOTERS

Thus far, this discussion has been restricted to promoters of sigma 70, the principal sigma factor of *E. coli* and related bacteria. There are a number of other sigma factors that direct the recognition of promoters with different DNA sequences (50, 67). These minor sigma factors fall into two distinct classes. Most of them are very similar to sigma 70 and differ largely in that small regions of the proteins are deleted. One expects that these will behave similarly to

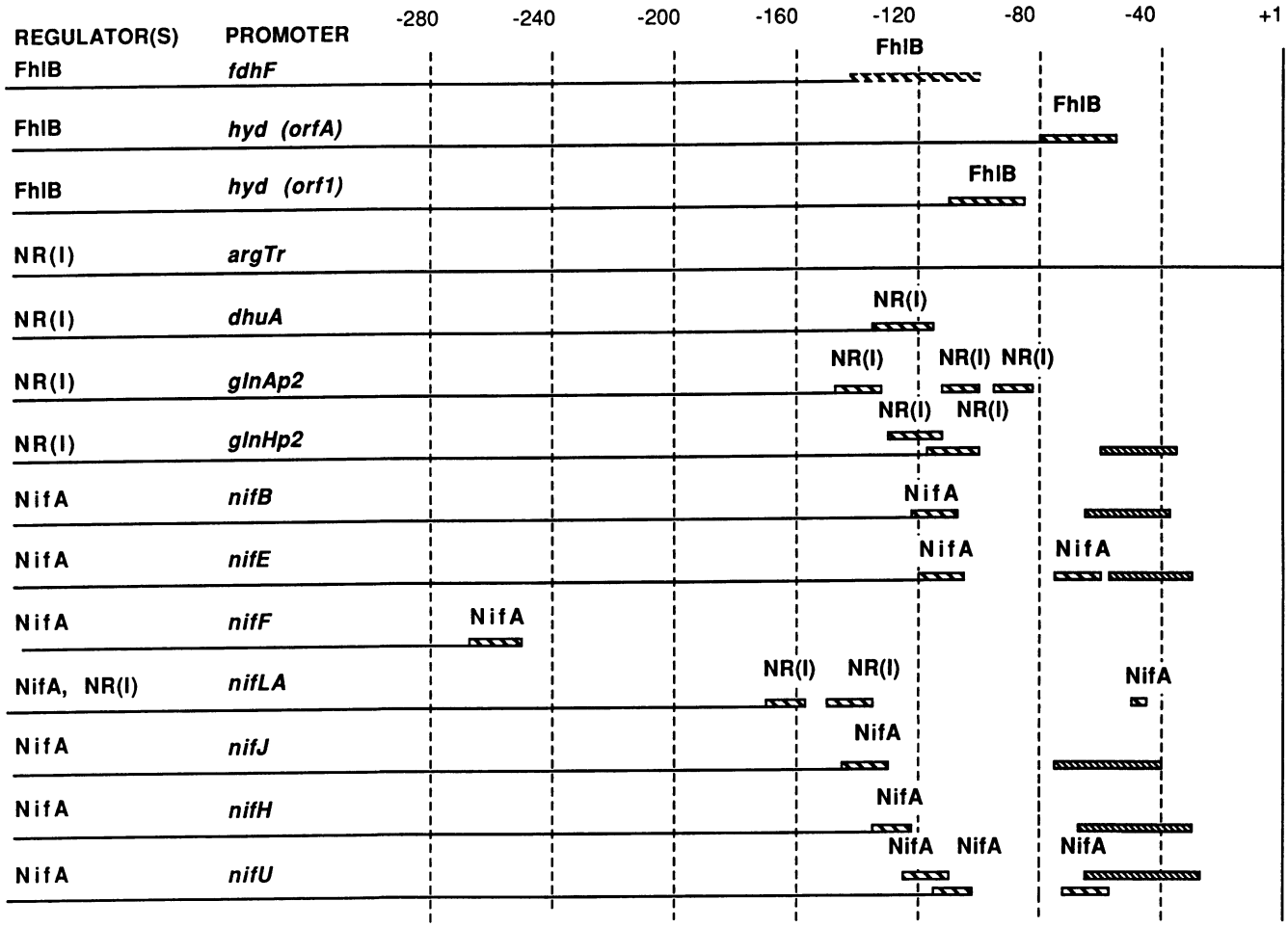

FIG. 3. Sigma 54 promoters. See legend to Fig. 1 for explanation of the order of promoters. Symbols: ▭, binding sites of activators; ▭, IHF-binding sites or a domain of activator-binding site defined by deletion analysis. References are indicated in Table 6.

TABLE 6. Sigma 54 promoters

Activator	Promoter	Type[a]	Evidence[b]	References
E. coli and *S. typhimurium*				
FhlB	*fdhF*	0	1	8, 9, 136
FhlB	*hyd-orfA*	D	4	79
FhlB	*hyd-orf1*	D	4	79
NR(I)	*argTr*	0	1	137
NR(I)	*dhuA*	0	2	3, 51
NR(I)	*glnAp2*	P	1,2**	67, 100
NR(I)	*glnHp2*	P	1,2	21a
K. pneumoniae				
NifA	*nifB*	0	4	5
NifA	*nifE*	0	4	5
NifA	*nifE*	D	1,4	93
NR(I) (NifA)	*nifLA*	D	1*	93
NifA	*nifH*	D	1*	13, 95, 129
NifA	*nifJ*	D	4*	5
NifA	*nifU*	0	1*	17

[a] Type refers to either sample promoters (0) or multiple promoters in a parallel (P) or divergent (D) array.

[b] Numbers indicating the type of evidence follow the same meaning as in Table 1. Symbols: *, in vivo footprinting; **, present in *S. typhimurium* with very similar organization (54).

sigma 70, and in the few cases studied this seems to be true (50). One sigma factor, sigma 54, is not at all similar to the others and is not considered part of the sigma 70 family of factors (67, 91). We now discuss how the arrangement of elements associated with sigma 54 promoters differs from that of sigma 70 promoters.

Arrangements of Regulatory Elements

Figure 3 and Table 6 collect the seven known examples of sigma 54 promoters from the same bacteria used to collect the sigma 70 data base, *E. coli* and *S. typhimurium*. Four of these are activated by the NR(I) (NtrC) activator and three are activated by the FhlB activator. Also collected in Fig. 3 are seven sigma 54 promoters from *Klebsiella pneumoniae*, all regulated by NifA. The sequence of the sigma 54 protein from this organism is almost identical to that from *E. coli*. These promoters are all activated by the NifA protein except *nifLA*, which is activated by NR(I). The three activators, NifA, NR(I), and FhlB, have certain structural similarities in the carboxyl DNA-binding end and in the central domain (31, 136).

Even though the number of examples is limited, it is clear that the organization of regulatory elements in sigma 54 promoters is quite different from that in sigma 70 promoters

in the same organisms [compare Fig. 1 and FhlB and NR(I) promoters in Fig. 3]. Whereas virtually all sigma 70 promoters contain proximal elements, six of these seven sigma 54 promoters do not contain them; the sole exception is the marginally proximal *hyd (orfA)* case, which is supported solely by DNA sequence similarity with another FhlB site. All of the NR(I) and two of three FhlB sites reside upstream from −85, and as a group they are centered near position −110. These are too remote to touch the polymerase without DNA looping, and, indeed, in vivo footprinting studies at *glnA* have shown a long stretch of unbound DNA between the NR(I) sites and the RNA polymerase recognition elements (130). Thus, sigma 54 promoters, in contrast to the basic sigma 70 promoter, can be activated without the need for proximal regulatory element to mediate the interaction with the RNA polymerase.

A slight modification of this interpretation may be necessary for the *glnH p2* promoter of *E. coli*, which uses the integration host factor (IHF) protein as a proximal cofactor (21a), as has been observed previously in several *nif* promoters of *K. pneumoniae* and other organisms (17, 57). In all of these sigma 54 promoters, IHF is not a traditional regulator since it does not stimulate transcription by itself. Rather, it seems to enhance the ability of remotely bound factors to activate (see below). Although evidence is lacking, it is conceivable that IHF could bind A+T-rich proximal sequences in other *E. coli* promoters such as *fdhF* and *hyd*.

This analysis can be extended to the seven known *Klebsiella* sigma 54 promoters. All seven *Klebsiella* promoters contain remote sites for the NifA or NR(I) activators, again centered far upstream, roughly near −120. Two promoters, *nifE* and *nifU*, contain additional putative sites just outside the proximal region near −70. The *nifLA* promoter is activated mainly by NR(I), which binds remotely, but can also be activated to a lesser extent by NifA, which binds a downstream site. In addition, five of these promoters contain a binding site for IHF.

Collectively, the sigma 54 data contrast very strongly with the arrangement of regulatory elements in the sigma 70 analysis. All fourteen sigma 54 promoters contain remote sites, compared with fewer than 10% of sigma 70 promoters. All of the sigma 54 promoters are subject to activation, and none is subject to repression, again in strong contrast to the more even distribution of control in sigma 70 promoters (Tables 1 and 2). As just discussed, only sigma 54 promoters do not require proximal sites for the binding of regulators. These considerations indicate that sigma 54 promoters as a class differ from sigma 70 promoters in that they are designed to be tightly regulated by activation alone and that this is accomplished primarily by a remote mechanism that is not easily accessible to the sigma 70 transcription apparatus.

These interpretations are supported by studies showing that promoter mutations have quite different consequences in the two classes of promoters. Up mutations in a weak sigma 70 promoter can enable it to have high transcriptional activity in the absence of activators. In fact, sigma 70 promoters can have significant intrinsic promoter activity that is further increased by activators (15, 60). By contrast, up mutations in sigma 54 promoters do not completely overcome the requirement for activators (but see reference 12). Moreover, there is no evidence that sigma 54 promoters have significant activity in the absence of activators. These differences confirm the nearly absolute dependence of sigma 54 promoters on activators, as suggested by the analysis of the data base.

The source of this difference may lie in the capability of sigma 54 polymerase to form stable, but inactive, closed complexes in the absence of activators (100, 110, 118, 131). If the bound sigma 54 polymerase is simply incapable of DNA melting, the requirement for activator in the melting step is exceptionally strong. By contrast, bound sigma 70 polymerase is clearly capable of melting the promoter DNA without the assistance of activators (44). Therefore, proximal promoter mutations can substitute for activators and assist the sigma 70 polymerase in both binding and melting, accounting for the widespread existence of activator-independent promoter mutations. By contrast, mutations in sigma 54 promoters would affect primarily the stability of the closed complex but could not direct the polymerase to melt the DNA in the absence of activator. This has been confirmed experimentally for the *nifH* promoter (12).

The important distinction of stable binding of inactive sigma 54 polymerase is best characterized in the *glnAp2* promoter of *S. typhimurium* and *E. coli* (100, 110, 118, 131) and has been reported also at *glnHp2* in *E. coli*, a mutation of *nifH* of *K. pneumoniae*, and *nifH* of *Rhizobium meliloti* (12, 21a). In the large sigma 70 promoter data base, only the *lac* promoter shows clear evidence of allowing stable binding by an inactive polymerase (see above). However, this is not an intrinsic property of the polymerase since it requires bound repressor to accomplish the stable binding in an inactive state. These comparisons support the idea that in the absence of effectors, bound sigma 70 polymerase is capable of DNA melting but bound sigma 54 polymerase is not. In addition, the paucity of evidence in the much larger group of sigma 70 promoters, as well as positive evidence in 5 of the 14 sigma 54 promoters, strongly indicates that this is a distinctive property of the type of promoter.

The exceptionally strong requirement for activators may simply make negative regulation by repressors unnecessarily redundant for sigma 54 promoters; recall that the data base with 14 cases shows no examples of repression, which is quite common for sigma 70 promoters. The idea that repression is unnecessary rather than impossible is supported by the fact that the important sigma 54 activator NR(I) acts as a repressor for the sigma 70 promoter *glnAp1* (118). That is, the protein has the ability to act as a repressor, but this is not used in the context of sigma 54 repression. Theoretically, there is no reason that sigma 54 transcription should be incompatible with repression, and perhaps such examples will emerge as the data base expands. By analogy with sigma 70 promoters, operators could overlap the activation site or the proximal elements.

Sigma 54 and sigma 70 bind the same core polymerase, and so the striking differences in control mechanism cannot be due to the core. Three different sigma 54 activators, NR(I), FhlB, and NifA, all allow activation without proximal elements (Fig. 3), a situation virtually forbidden at sigma 70 promoters. These activator proteins have some similarity, as already mentioned, but are for the most part different and function in distinct regulatory pathways. One of these sigma 54 activators, NR(I), fails to activate the sigma 70-dependent *lac* promoter when a site for it replaces that of the sigma 70 catabolite activator protein and is located at a similar distance from the position where it usually activates *glnAp2* (117). Therefore, the most likely determinant of the common ability of these systems to work without proximal activator proteins is the common involvement of sigma 54 rather than sigma 70.

Since sigma 54 mediates activation without accessory proximal elements, one might expect that the sigma 54 protein itself would somehow substitute for this function of

the proximal sigma 70 activation elements. The structure of the sigma 54 protein appears to allow this by containing domains that allow stable promoter recognition while retaining responsiveness to physiological activation after stable binding has occurred (132). This domain structure allows it to catalyze a mechanism of activation that is strongly disfavored in promoters transcribed by the sigma 70 family of factors, which have a quite different domain structure.

Recall that in many cases the purpose of proximal elements in sigma 70 promoters was to allow the RNA polymerase to be touched in order to enhance its binding, which is intrinsically defective as a result of the lack of optimal −35 recognition elements. Stabilizing interactions diminish with distance, accounting for the close proximity of the sigma 70 activators (see above). Since sigma 54 polymerase binds tightly to *glnAp2* without activator, the deleterious effect of distance on cooperative binding is less important. The activator must merely trigger a conformational change in a prebound complex, accomplishing only the second half of what the sigma 70 activators do from a nearby position.

This mechanistic difference probably also contributes to the unusual flexibility in the positioning of the *E. coli* sigma 54-dependent activation sites. The data base already hints at this flexibility, as the sigma 54 activator sites in Fig. 3 are much more dispersed than those for sigma 70 in Fig. 2. Recall that for the sigma 70 activators OmpR and CRP, activation diminishes precipitously as the DNA elements are moved out of the nearby proximal position (40, 81). By contrast, the NR(I)-binding sites of *glnAp2* can be moved more than 1 kb upstream or downstream and still retain 50% of their function (120). The FhlaB site can be moved 2 kb from the *fdhF* promoter and still retains 20% of its function (8), and at the same distance, the NifA site still retains 10% of activation on *nifH* (13). Since sigma 54 itself is making an important contribution to promoter recognition, the NR(I), NifA, and FhlB factors need to contribute much less to promoter recognition by the polymerase and thus are less strongly affected by their removal to long distances.

The mechanism of activation by the sigma 54 polymerase is similar to activation by sigma 70 polymerase in one sense: both appear to involve the activator's touching the RNA polymerase at the promoter. For sigma 54, the activator sites are nonproximal and the mechanism has been shown to involve looping out of the intervening DNA (148). A different view of the role of stable closed-complex formation in this mechanism can be illustrated by comparing hypothetical long-range activation at the *lac* and *gln* promoters. If the *lac* CRP site were far away, most looping events would bring CRP to an unoccupied promoter since sigma 70 polymerase is unable to recognize the *lac* promoter on its own (175). Activation could only occur in the unlikely circumstance that transient promoter recognition had occurred at the precise moment when CRP was brought nearby by DNA looping. At the *glnAp2* promoter, by contrast, when remotely bound NR(I) is brought near the promoter by DNA looping, it will inevitably encounter RNA polymerase that has been directed to bind by sigma 54. Thus, the probability of the remotely bound activator touching the polymerase at the promoter is much greater for *gln* than for *lac*, accounting for the difference in their abilities to be activated from great distances.

These observations provide a rationalization for the differing organizations of the sigma 54 promoters in Fig. 3 and the sigma 70 promoters in Fig. 1. All the sigma 54 promoters require activation since the stable closed complex is unable to melt the DNA on its own. The position of the sigma 54 activator sites in Fig. 3 is much less fixed than that observed for sigma 70 activators such as FNR or CRP (see above). This is easily rationalized in terms of using a DNA-looping mechanism rather than a stereospecific mechanism as proposed for sigma 70. There may be an additional contribution to sigma 54 activation, as suggested by studies on *argTr* and *glnAp2*. At the *argTr* promoter, activation appears to occur without either proximal or remote sites. Despite the lack of an activator-binding site, this sigma 54 promoter is activated by NR(I) (137), although one cannot rule out a contribution to the activation of this promoter in vivo by the NR(I) site of *dhuA*, which is located less than 1 kb downstream (3, 51). At *glnAp2* the dependence of activation on the NR(I) DNA-binding site lessens as the concentration of NR(I) increases (120). Thus, the consequences for activation of moving the activator sites to far distances may be minimal for these systems, as larger amounts of protein can partially compensate for unfavorable arrangements. This may also contribute to the remote and variable distribution of activator sites for sigma 54 promoters.

Proximal Coactivation by IHF

This discussion has emphasized the variable positioning and the nonproximal location of the sigma 54 activation sites. The locations have been rationalized in terms of the known involvement of a stable closed proximal complex and DNA looping to deliver the activator to the proximal polymerase. However, some of these promoters contain a proximal IHF-binding site, generally near −40 or −50. As mentioned above, IHF by itself does not stimulate transcription in these promoters, but enhances the effectiveness of a required activator. The potential involvement of IHF will now be discussed.

When IHF sites appear, they are always located between the polymerase and activator elements, as would be required if they assisted in bringing the two together by enhancing loop formation (57). If IHF is to help the activation mechanism by bending the DNA, it is expected that it will be useful only when the activator sites are located such that the bending of the DNA will bring the bound activator into a position favorable for contact with the closed complex. The distance between the IHF site and the upstream activator site is about 80 bp in all cases (Fig. 3), suggesting that this distance is important. The mechanism is supported further by the conversion of IHF into a repressor of *glnHp2* when the distance between the NR(I)-binding site and the promoter is modified (21a). Such repressor effects have been observed in sigma 70 promoters with IHF-binding sites, although the mechanisms are still unknown (58, 64).

IHF assistance in DNA looping may compensate for the low intrinsic stability of the looped nucleoprotein complex. In two *Klebsiella* promoters with IHF sites, the *nifH* and *nifLA* promoters, the sigma 54 polymerase cannot accomplish fully stable promoter recognition in the absence of activator. This is apparently related to the DNA sequences of the promoters themselves rather than being a property of sigma 54; a point mutation in the *nifH* DNA elements recognized by sigma 54 polymerase increases transcriptional activity (12), and a 3-base C-to-T substitution allows stable closed promoter recognition to occur (96). This suggests that the *Klebsiella* sigma 54 polymerase makes a very substantial contribution to stabilizing closed-complex formation but not enough to direct full occupancy of the *nifH* promoter. Thus, in at least two cases, the arrangement and DNA sequence of elements in the *Klebsiella* sigma 54 promoters are such that

full advantage is not taken of the ability of the polymerase to form a stable closed complex. This means that when the bound activator loops to the promoter, it is less likely to encounter a bound sigma 54 polymerase. In apparent compensation for this, the bending protein IHF has a binding site within the loop and promotes DNA bending so as to increase the probability of loop formation. Thus, the bound activator is held near the promoter for a longer time, allowing a greater probability of touching the polymerase which is bound in a short-lived complex. In support of this idea, much less assistance is given by IHF in the mutant *nifH* promoter that supports stable closed-complex formation (57).

The role of IHF need not be restricted to assisting in loop formation when closed complexes are unstable. In principle, bound IHF could enhance any step leading to a final complex in which bound polymerase and bound activator touch. Examples might include enhancing the binding of a weakly bound activator or enhancing the contact between activators and sigma 54 polymerase that bind each other weakly. In these cases, bending will assist in complex formation, thereby enhancing the binding of all components. One example may be the involvement of an IHF site in *glnHp2*, a strong promoter able to bind RNA polymerase in an inactive stable complex. It has been argued that IHF compensates for the weakened binding of NR(I) since overlapping sites allow only one site to be available for binding (21a). In addition, since IHF sites are located just adjacent to the polymerase-binding site, a contribution of activation by direct touching of the polymerase or the activator is possible.

Thus, it appears that some sigma 54 promoters have evolved an organization in which suboptimal sigma 54 recognition sites, as illustrated by *nifH*, or suboptimal activation sites, as illustrated by *glnHp2*, are coupled with nearby compensating IHF sites. The current state of knowledge about these and other IHF-containing promoters is not sufficient for us to know how well these limited examples cover the range of IHF assistance. For example, in the *glnHp2* promoter IHF does not compensate for low RNA polymerase-sigma 54 concentrations, indicating that at least in some cases, IHF is not able to compensate for a poor promoter recognition.

The *glnAp2* promoter, which has no IHF-binding sites, can be 50% activated by NR(I) sites placed 2 kb away, but the *nifH* promoter, which has an IHF-binding site, retains only 10% activation by NifA sites. In general the IHF sites cataloged in Fig. 3 appear to occur in poor promoters, as evaluated by the deviation of promoter sequences from consensus. It is not yet known whether the difference in strength of remote activation is due to differences in intrinsic promoter strength or the differential involvement of IHF. In general, such differences could be due to the differing role of IHF as a coactivator or repressor (see above), the use of different activators, or differences in the stability of closed complexes.

Sigma 54 Promoter Organization and Mechanism: an Overview

One further requirement for the arrangement of sigma 54 promoters is that the promoters be designed so that they are very well separated or have another means of avoiding inappropriate cross-activation. This appears not to be a significant problem for sigma 54 promoters of *E. coli* since these promoters are rare. Thus, even though the activator sites can work over long distances, there are no other sigma 54 promoters within range. Such is not the case, however, for the divergently oriented *hyd* promoters of *E. coli*, the divergent *nifF* and *nifLA* promoters of *K. pneumoniae*, and the promoters of histidine transport, *argTr* and *dhuA*. Of these three pairs, two contain a single IHF site, implying that there is no general requirement for proximal IHF sites in closely spaced sigma 54 promoters. However, the problem of inappropriate cross-activation does not even arise in these cases, since each promoter pair is part of the same physiological response system.

It is interesting that the sigma 54 mechanism is more flexible than the sigma 70 mechanism but requires more DNA. The compactness of the *E. coli* genome is compatible with the more highly evolved sigma 70 system which dominates its genetic organization. The potential flexibility in regulation offered by systems like those of sigma 54 promoters would be better used in organisms that can tolerate larger amounts of DNA. This is an unlikely alternative for *E. coli*, which must compete for rapid growth in energy-limited environments (82).

In summary, the sigma 54 promoters as a class differ in the arrangements of regulatory sites from the sigma 70 family of promoters. These differences in arrangement can be rationalized in terms of different mechanisms of control of holoenzymes containing different sigma factors. For sigma 70, proximal elements are used to mediate stereospecific communication with the polymerase, which generally needs assistance in binding to promoters with poor recognition elements. For sigma 54, DNA looping from variable distances is used to touch the polymerase, which has substantial ability to bind on its own. The latter mechanism seems to be sufficient for regulation, since simple repression using operators appears not to occur.

In these regards, the sigma 54 promoters resemble eukaryotic promoters closely, whereas the sigma 70 promoters do not. Prokaryotic sigma 54 promoters represent, at least conceptually, an intermediate step in the history of regulation between prokaryotic and eukaryotic cells. This is supported by the following set of interrelated common properties for sigma 54 and eukaryotic promoters: (i) long-distance activation associated with a requirement for large amounts of DNA; (ii) stable preinitiation complexes (14, 172); (iii) a requirement for ATP in proximal assembly; (iv) predominant occurrence of activation versus simple repression regulation associated with a low basal level of transcription in the absence of activators (73, 74, 112) (Fig. 1); and (v) structural similarity of domains present in sigma 54 with motifs present in eukaryotic activators and absent in other sigma factors (132). The interdependence of these properties as a consequence of the sigma 54 activation mechanism has been discussed throughout the review.

The evolutionary distance between the origin of bacteria and that of the simplest eukaryotic cells is very large. The difficulties in establishing a bridge between these two worlds are illustrated by comparing sigma 54 and yeast TFIID, the DNA-binding factor required for transcription of eukaryotic RNA polymerase II. The similarity of these two proteins is not greater than that of TFIID and other sigma factors (unpublished observation). On the other hand, sigma 54 shares with the TFIID molecules of *Drosophila melanogaster* and humans, but not with the one of yeasts, the occurrence of domains rich in glutamine residues, as well as other domains also present in eukaryotic activators (98, 132). Deciphering this puzzle of molecular similarities could clarify the origin of eukaryotic promoters and define whether the sigma 54 bacterial promoters were evolutionary precursors.

Considering the collection of properties associated with the regulation of sigma 54 promoters, it is clear that, in any case, these promoters were at least a good experiment in the evolution of what later became important properties of gene regulation in eukaryotes. These considerations should provide food for thought to geneticists, biochemists, biologists, and evolutionists.

ACKNOWLEDGMENTS

J.C.-V. has been supported by a postdoctoral fellowship from the Universidad Nacional Autónoma de México (UNAM) and more recently by Public Health Service Fogarty International Research Fellowship no. F05-TW04437. This work was initiated while J.D.G. was on sabbatical leave at MIT. Preparation of this article was supported by Public Health Service grant GM35754 to J.D.G.

We thank Bo Tendis for her help in the preparation of the manuscript.

ADDENDUM IN PROOF

Ostrovsky de Spicer et al. have recently shown (P. Ostrovsky de Spicer, K. O'Brien, and S. Maloy, J. Bacteriol. **173**:211–219, 1991) that *putA* is not an exception, as it uses a remote duplicated operator.

REFERENCES

1. **Adhya, S.** 1987. The galactose operon, p. 1503–1512. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
2. **Aiba, H.** 1985. Transcription of the *Escherichia coli* adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. J. Biol. Chem. **260**:3063–3070.
3. **Ames, G. F.-L., and K. Nikaido.** 1985. Nitrogen regulation in *Salmonella typhimurium*. Identification of an *ntrC* protein-binding site and definition of a consensus binding sequence. EMBO J. **4**:539–547.
4. **Amouyal, M., L. Mortensen, H. Buc, and K. Hammer.** 1988. Single and double loop formation when *deoR* repressor binds to its natural operator sites. Cell **58**:545–551.
5. **Arnold, W., A. Rump, W. Klipp, U. B. Priefer, and A. Pühler.** 1988. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. J. Mol. Biol. **203**:715–738.
6. **Aslanidis, C., and R. Schmitt.** 1990. Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. J. Bacteriol. **172**:2178–2180.
7. **Baseggio, N., W. D. Davies, and B. E. Davidson.** 1990. Identification of the promoter, operator, and 5′ and 3′ ends of the mRNA of the *Escherichia coli* K-12 gene *aroG*. J. Bacteriol. **172**:2547–2557.
8. **Birkmann, A., and A. Böck.** 1989. Characterization of a *cis* regulatory DNA element is necessary for formate induction of the formate dehydrogenase gene (*fdhF*) of *Escherichia coli*. Mol. Microbiol. **3**:187–195.
9. **Birkmann, A., F. Zinoni, G. Sawers, and A. Böck.** 1987. Factors affecting transcriptional regulation of the formate-hydrogen-lyase pathway of *Escherichia coli*. Arch. Microbiol. **148**:44–51.
10. **Brandsma, J. A., D. Bosch, M. de Ruÿter, and P. van de Putte.** 1985. Analysis of the regulatory region of the *ssb* gene of *Escherichia coli*. Nucleic Acids Res. **13**:5095–5109.
11. **Brent, R., and M. Ptashne.** 1981. Mechanism of action of the *lexA* gene product. Proc. Natl. Acad. Sci. USA **78**:4204–4208.
12. **Buck, M., and W. Cannon.** 1989. Mutations in the RNA polymerase recognition sequence of the *Klebsiella pneumoniae nifH* promoter permitting transcriptional activation in the absence of NifA binding to upstream activator sequences. Nucleic Acids Res. **17**:2597–2612.
13. **Buck, M., S. Miller, M. Drummond, and R. Dixon.** 1986. Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature (London) **320**:374–378.
14. **Buratowski, S., S. Hahn, L. Guarente, and P. A. Sharp.** 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell **56**:549–561.
15. **Busby, S. J. W.** 1986. Positive regulation in gene expression, p. 51–77. *In* I. R. Booth and C. F. Higgins (ed.), Regulation of gene expression—25 years on. Cambridge University Press, Cambridge, United Kingdom.
16. **Calderwood, S. B., and J. J. Mekalanos.** 1988. Confirmation of the Fur operator site by insertion of a synthetic oligonucleotide into an operon fusion plasmid. J. Bacteriol. **170**:1015–1017.
17. **Cannon, W. V., R. Kreutzer, H. M. Kent, E. Morett, and M. Buck.** 1990. Activation of the *Klebsiella pneumoniae nifU* promoter: identification of multiple and overlapping upstream NifA binding sites. Nucleic Acids Res. **18**:1693–1701.
18. **Charlier, D., G. Weyens, M. Roovers, J. Piette, C. Bocquet, A. Piérard, and N. Glansdorff.** 1988. Molecular interactions in the control region of the *carAB* operon encoding *Escherichia coli* carbamoylphosphate synthetase. J. Mol. Biol. **204**:867–877.
19. **Choi, K. Y., and H. Zalkin.** 1990. Regulation of *Escherichia coli pyrC* by the purine regulon repressor protein. J. Bacteriol. **172**:3201–3207.
20. **Christman, M. F., G. Storz, and B. N. Ames.** 1989. OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulator proteins. Proc. Natl. Acad. Sci. USA **86**:3484–3488.
21. **Chye, M.-L., and J. Pittard.** 1987. Transcription control of the *aroP* gene in *Escherichia coli* K-12: analysis of operator mutants. J. Bacteriol. **169**:386–393.

21a. **Claverie-Martin, F., and B. Magasanik.** 1991. Role of integration host factor in the regulation of the *glnHp2* promoter of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88**:1631–1635.

22. **Cobbett, C. S.** 1988. Repression of the *aroF* promoter by the TyrR repressor in *Escherichia coli* K-12: role of the 'upstream' operator site. Mol. Microbiol. **2**:377–383.
23. **Coulton, J. W., P. Mason, D. R. Cameron, G. Carmel, R. Jean, and H. N. Rode.** 1986. Protein fusions of β-galactosidase to the ferrichrome-iron receptor of *Escherichia coli* K-12. J. Bacteriol. **165**:181–192.
24. **Cronan, J. E., Jr.** 1989. The *E. coli bio* operon: transcriptional repression by an essential protein modification enzyme. Cell **58**:427–429.
25. **Cunin, R., T. Eckhardt, J. Piette, A. Boyen, A. Piérard, and N. Glansdorff.** 1983. Molecular basis for modulated regulation of gene expression in the arginine regulon of *Escherichia coli* K-12. Nucleic Acids Res. **11**:5007–5019.
26. **de Crombrugghe, B., S. Busby, and H. Buc.** 1984. Activation of transcription by the cyclic AMP receptor protein, p. 129–167. *In* K. Yamamoto (ed.), Biological regulation and development, vol. III-B. Plenum Press, New York.
27. **Deeley, M. C., and C. Yanofsky.** 1982. Transcription initiation at the tryptophanase promoter of *Escherichia coli* K-12. J. Bacteriol. **151**:942–951.
28. **DeFeyter, R. C., B. E. Davidson, and J. Pittard.** 1986. Nucleotide sequence of transcription unit containing the *aroL* and *aroM* genes from *Escherichia coli* K-12. J. Bacteriol. **165**:233–239.
29. **de Lorenzo, V., F. Giovannini, M. Herrero, and J. B. Neilands.** 1988. Metal ion regulation of gene expression. Fur repressor-operator interaction at the promoter region of the aerobactin system of pColV-K30. J. Mol. Biol. **203**:875–884.
30. **de Lorenzo, V., S. Wee, M. Herrero, and J. B. Neilands.** 1987. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. J. Bacteriol. **169**:2624–2630.
31. **Drummond, M., P. Whitty, and J. Wootton.** 1986. Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*: homologies to other regulatory proteins. EMBO J. **5**:441–447.
32. **Easton, A. M., and S. R. Kushner.** 1983. Transcription of the *uvrD* gene of *Escherichia coli* is controlled by the *lexA* repres-

sor and by attenuation. Nucleic Acids Res. **11**:8625–8640.
33. **Ebina, Y., F. Kishi, T. Miki, H. Kagamiyama, T. Nakazawa, and A. Nakazawa.** 1981. The nucleotide sequence surrounding the promoter region of colicin E1 gene. Gene **15**:119–126.
34. **Ebina, Y., Y. Takahara, F. Kishi, and A. Nakazawa.** 1983. LexA protein is a repressor of the colicin E1 gene. J. Biol. Chem. **258**:13258–13261.
35. **Eiglmeier, K., N. Honoré, S. Iuchi, E. C. C. Lin, and S. T. Cole.** 1989. Molecular genetic analysis of FNR-dependent promoters. Mol. Microbiol. **3**:869–878.
36. **Flashner, Y., and J. D. Gralla.** 1988. Dual mechanism of repression at a distance in the *lac* operon. Proc. Natl. Acad. Sci. USA **85**:8968–8972.
37. **Forsman, K., M. Göransson, and B. E. Uhlin.** 1989. Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. EMBO J. **8**:1271–1277.
38. **Forst, S. A., J. Delgado, and M. Inouye.** 1989. DNA-binding properties of the transcription activator (OmpR) for the upstream sequences of *ompF* in *Escherichia coli* are altered by *envZ* mutations and medium osmolarity. J. Bacteriol. **171**: 2949–2955.
39. **Friden, P., T. Newman, and M. Freundlich.** 1982. Nucleotide sequence of the ilvB promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. Proc. Natl. Acad. Sci. USA **79**:6156–6160.
40. **Gaston, K., A. Bell, A. Kolb, H. Buc, and S. Busby.** 1990. Stringent spacing requirements for transcription activation by CRP. Cell **62**:733–743.
41. **Göransson, M., K. Forsman, P. Nilsson, and B. E. Uhlin.** 1989. Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in *Escherichia coli*. Mol. Microbiol. **3**:1557–1565.
42. **Gralla, J. D.** 1989. Bacterial gene regulation from distant DNA sites. Cell **57**:193–195.
43. **Gralla, J. D.** 1989. Specific repression in the *lac* operon, p. 3–10. *In* J. D. Gralla (ed.), DNA-protein interactions in transcription. Alan R. Liss, Inc., New York.
44. **Gralla, J. D.** 1990. Promoter recognition and mRNA initiation by *Escherichia coli* E-sigma-70. Methods Enzymol. **185**:37–54.
45. **Griggs, D. W., and J. Konisky.** 1989. Mechanism for iron-regulated transcription of the *Escherichia coli cir* gene: metal-dependent binding of Fur protein to the promoters. J. Bacteriol. **171**:1048–1054.
46. **Haber, R., and S. Adhya.** 1988. Interaction of spatially separated protein-DNA complexes for control of gene expression: operator conversions. Proc. Natl. Acad. Sci. USA **85**:9683–9687.
47. **Hahn, D. R., R. S. Myers, C. R. Kent, and S. R. Maloy.** 1988. Regulation of proline utilization in *Salmonella typhimurium*: molecular characterization of the *put* operon, and DNA sequence of the *put* control region. Mol. Gen. Genet. **213**:125–133.
48. **Hamilton, E. P., and N. Lee.** 1988. Three binding sites for AraC protein are required for autoregulation of *araC* in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:1749–1753.
49. **He, B., A. Shiau, K. Y. Choi, H. Zalkin, and J. M. Smith.** 1990. Genes of the *Escherichia coli pur* regulon are negatively controlled by a repressor-operator interaction. J. Bacteriol. **172**:4555–4562.
50. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57**:839–872.
51. **Higgins, C. F., and G. F.-L. Ames.** 1982. Regulatory regions of two transport operons under nitrogen control: nucleotide sequences. Proc. Natl. Acad. Sci. USA **79**:1083–1087.
52. **Hillen, W., C. Gatz, L. Altschmied, K. Schollmeier, and I. Meier.** 1983. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. J. Mol. Biol. **169**:707–721.
53. **Hillen, W., K. Schollmeier, and C. Gatz.** 1984. Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the *tet* operon regulatory region. J. Mol. Biol. **172**:185–201.
54. **Hirschman, J., P. K. Wong, S. J. Keener, and S. Kustu.** 1985. Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription *in vitro*: evidence that the *ntrA* product is a sigma factor. Proc. Natl. Acad. Sci. USA **82**:7525–7529.
55. **Ho, Y.-S., D. L. Wulff, and M. Rosenberg.** 1983. Bacteriophage lambda protein cII binds promoters on the opposite face of the DNA helix from RNA polymerase. Nature (London) **304**:703–708.
56. **Hochschild, A., and M. Ptashne.** 1986. Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. Cell **44**:681–687.
57. **Hoover, T. R., E. Santero, S. Porter, and S. Kustu.** 1990. The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. Cell **63**:11–22.
58. **Huang, L., P. Tsui, and M. Freundlich.** 1990. Integration host factor is a negative effector of in vivo and in vitro expression of *ompC* in *Escherichia coli*. J. Bacteriol. **172**:5293–5298.
59. **Huo, L., K. J. Martin, and R. Schleif.** 1988. Alternative DNA loops regulate the arabinose operon in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:5444–5448.
60. **Irani, M., R. Musso, and S. Adhya.** 1989. Cyclic-AMP-dependent switch in initiation of transcription from the two promoters of the *Escherichia coli gal* operon: identification and assay of 5′-triphosphate ends of mRNA by GTP:RNA guanyltransferase. J. Bacteriol. **171**:1623–1630.
61. **Jayaraman, P. S., J. A. Cole, and S. J. W. Busby.** 1989. Mutational analysis of the nucleotide sequence at the FNR-dependent *nirB* promoter in *Escherichia coli*. Nucleic Acids Res. **17**:135–145.
62. **Jayaraman, P.-S., K. L. Gaston, J. A. Cole, and S. J. W. Busby.** 1988. The *nirB* promoter of *Escherichia coli*: location of nucleotide sequences essential for regulation by oxygen, the FNR protein and nitrite. Mol. Microbiol. **2**:527–530.
63. **Kasian, P. A., B. E. Davidson, and J. Pittard.** 1986. Molecular analysis of the promoter operator region of the *Escherichia coli* K-12 *tyrP* gene. J. Bacteriol. **167**:556–561.
64. **Kennedy, M., M. Chandler, and D. Lane.** 1988. Mapping and regulation of the *pifC* promoter of the F plasmid. Biochim. Biophys. Acta **950**:75–80.
65. **Klig, L. S., J. Carey, and C. Yanofsky.** 1988. *trp* repressor interactions with the *trp aroH* and *trpR* operators. Comparison of repressor binding *in vitro* and repression *in vivo*. J. Mol. Biol. **202**:769–777.
66. **Kuhnke, G., C. Theres, H. J. Fritz, and R. Ehring.** 1989. RNA polymerase and *gal* repressor bind simultaneously and with DNA bending to the control region of the *Escherichia coli* galactose operon. EMBO J. **8**:1247–1255.
67. **Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss.** 1989. Expression of sigma 54 (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol. Rev. **53**:367–376.
68. **Lanzer, M., and H. Bujard.** 1988. Promoters largely determine the efficiency of repressor action. Proc. Natl. Acad. Sci. USA **85**:8973–8977.
69. **Lee, D. H., and R. F. Schleif.** 1989. *In vivo* DNA loops in *araCBAD*: size limits and helical repeat. Proc. Natl. Acad. Sci. USA **86**:476–480.
70. **Lee, N. L., C. Franklyn, and E. P. Hamilton.** 1987. Arabinose-induced binding of AraC protein to *araI2* activates the *araBAD* operon promoter. Proc. Natl. Acad. Sci. USA **84**:8814–8818.
71. **Lee, N. L., W. O. Gielow, and R. G. Wallace.** 1981. Mechanism of *araC* autoregulation and the domains of two operlapping promoters, PC and PBAD, in the L-arabinose regulatory region of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:742–756.
72. **Le Grice, S. F. J., H. Matzura, R. Marcoli, S. Iida, and T. A. Bickle.** 1982. The catabolite-sensitive promoter for the chloramphenicol acetyltransferase gene is preceded by two binding sites for the catabolite gene activator protein. J. Bacteriol. **150**:312–318.

73. **Levine, M., and J. L. Manley.** 1989. Transcriptional repression of eukaryotic promoters. Cell **59**:405–408.
74. **Lewin, B.** 1990. Commitment and activation at Pol II promoters: a tail of protein-protein interactions. Cell **61**:1161–1164.
75. **Li, S. F., and J. A. DeMoss.** 1988. Location of sequences in the *nar* promoter of *Escherichia coli* required for regulation by Fnr and NarL. J. Biol. Chem. **263**:13700–13705.
76. **Lim, D., J. D. Oppenheim, T. Eckhardt, and W. K. Maas.** 1987. Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. Proc. Natl. Acad. Sci. USA **84**:6697–6701.
77. **Little, J. W., D. W. Mount, and C. R. Yanisch-Perron.** 1981. Purified *lexA* protein is a repressor of the *recA* and *lexA* genes. Proc. Natl. Acad. Sci. USA **78**:4199–4203.
78. **Lund, P. A., and N. L. Brown.** 1989. Regulation of transcription in *Escherichia coli* from the *mer* and *merR* promoters in the transposon Tn501. J. Mol. Biol. **205**:343–353.
79. **Lutz, S., R. Bohm, A. Beier, and A. Böck.** 1990. Characterization of divergent NtrA-dependent promoters in the anaerobically expressed gene cluster coding for hydrogenase 3 components of *Escherichia coli*. Mol. Microbiol. **4**:13–20.
80. **Maeda, S., and T. Mizuno.** 1990. Evidence for multiple OmpR-binding sites in the upstream activation sequence of the *ompC* promoter in *Escherichia coli*: a single OmpR-binding site is capable of activating the promoter. J. Bacteriol. **172**:501–503.
81. **Maeda, S., Y. Ozawa, T. Mizuno, and S. Mizushima.** 1988. Stereospecific positioning of the *cis*-acting sequence with respect to the canonical promoter is required for activation of the *ompC* gene by a positive regulator, OmpR, in *Escherichia coli*. J. Mol. Biol. **202**:433–441.
82. **Magasanik, B.** 1989. Gene regulation from sites near and far. New Biol. **1**:247–251.
83. **Makino, K., H. Shinagawa, M. Amemura, S. Kimura, and A. Nakata.** 1988. Regulation of the phosphate regulon of *Escherichia coli*. Activation of *pstS* transcription by PhoB protein *in vitro*. J. Mol. Biol. **203**:85–95.
84. **Makino, K., H. Shinagawa, M. Amemura, and A. Nakata.** 1986. Nucleotide sequence of the *phoB* gene, the positive regulatory gene for the phosphate regulon of *Escherichia coli* K-12. J. Mol. Biol. **190**:37–44.
85. **Mandal, N., W. Su, R. Haber, S. Adhya, and H. Echols.** 1990. DNA looping in cellular repression of transcription of the galactose operon. Genes Dev. **4**:410–418.
86. **Martinussen, J., N. E. Møllegaard, B. Holst, S. R. Douthwaite, and P. Valentin-Hansen.** 1989. A new version of negative control. DNA sequences involved in expression and regulation of CYTR and cAMP/CRP controlled genes in *Escherichia coli*, p. 31–41. *In* J. D. Gralla (ed.), DNA-protein interactions in transcription. Alan R. Liss, Inc., New York.
87. **Meiklejohn, A. L., and J. D. Gralla.** 1985. Entry of RNA polymerase at the *lac* promoter. Cell **43**:769–776.
88. **Menendez, M., A. Kolb, and H. Buc.** 1987. A new target for CRP action at the *malT* promoter. EMBO J. **6**:4227–4234.
89. **Meng, L. M., M. Kilstrup, and P. Nygaard.** 1990. Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN* and *guaBA* expression in *Escherichia coli*. Eur. J. Biochem. **187**:373–379.
90. **Menon, K. P., and N. L. Lee.** 1990. Activation of *ara* operons by a truncated AraC protein does not require inducer. Proc. Natl. Acad. Sci. USA **87**:3708–3712.
91. **Merrick, M., J. Gibbins, and A. Toukdarian.** 1987. The nucleotide sequence of the sigma factor gene *ntrA* (*rpoN*) of *Azotobacter vinelandii*: analysis of conserved sequences in NtrA proteins. Mol. Gen. Genet. **210**:323–330.
92. **Michaeli, S., M. Mevarech, and E. Z. Ron.** 1984. Regulatory region of the *metA* gene of *Escherichia coli* K-12. J. Bacteriol. **160**:1158–1162.
93. **Minchin, S. D., S. Austin, and R. A. Dixon.** 1988. The role of activator binding sites in transcriptional control of the divergently transcribed *nifF* and *nifLA* promoters from *Klebsiella pneumoniae*. Mol. Microbiol. **2**:433–442.
94. **Mizusawa, S., D. Court, and S. Gottesman.** 1983. Transcription of the *sulA* gene and repression by LexA. J. Mol. Biol. **171**:337–343.
95. **Morett, E., and M. Buck.** 1988. NifA-dependent *in vivo* protection demonstrates that the upstream activator sequence of *nif* promoters is a protein binding site. Proc. Natl. Acad. Sci. USA **85**:9401–9405.
96. **Morett, E., and M. Buck.** 1989. *In vivo* studies on the interaction of RNA polymerase sigma-54 with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. The role of NifA in the formation of an open promoter complex. J. Mol. Biol. **210**:65–77.
97. **Movva, R. N., P. Green, K. Nakamura, and M. Inouye.** 1981. Interaction of cAMP receptor protein with the *ompA* gene, a gene for a major outer membrane protein of *Escherichia coli*. FEBS Lett. **128**:186–190.
98. **Muhich, M. L., C. T. Iida, M. Horikoshi, R. G. Roeder, and C. S. Parker.** 1990. cDNA clone encoding *Drosophila* transcription factor TFIID. Proc. Natl. Acad. Sci. USA **87**:9148–9152.
99. **Nagai, H., R. Yano, J. E. Erickson, and T. Yura.** 1990. Transcriptional regulation of the heat shock regulatory gene *rpoH* in *Escherichia coli*: involvement of a novel catabolite-sensitive promoter. J. Bacteriol. **172**:2710–2715.
100. **Ninfa, A. J., L. J. Reitzer, and B. Magasanik.** 1987. Initiation of transcription at the bacterial *glnAp2* promoter by purified *E. coli* components is facilitated by enhancers. Cell **50**:1039–1046.
101. **Oehler, S., E. R. Eismann, H. Krämer, and B. Müller-Hill.** 1990. The three operators of the *lac* operon cooperate in repression. EMBO J. **9**:973–979.
102. **Ogden, S., D. Haggerty, C. M. Stoner, D. Kolodrubetz, and R. Schleif.** 1980. The *Escherichia coli* L-arabinose operon: binding of the regulatory proteins and a mechanism of positive and negative regulation. Proc. Natl. Acad. Sci. USA **77**:3346–3350.
103. **O'Halloran, T. V., B. Frantz, M. K. Shin, D. M. Ralston, and J. G. Wright.** 1989. The MerR heavy metal receptor mediates positive activation in a topologically novel transcription complex. Cell **56**:119–129.
104. **Otsuka, A., and J. Abelson.** 1978. The regulatory region of the biotin operon in *Escherichia coli*. Nature (London) **276**:689–694.
105. **Pinkney, M., and J. G. Hoggett.** 1988. Binding of the cyclic AMP receptor protein of *Escherichia coli* to RNA polymerase. Biochem. J. **250**:897–902.
106. **Plamann, L. S., and G. V. Stauffer.** 1987. Nucleotide sequence of the *Salmonella typhimurium metR* gene and the *metR-metE* control region. J. Bacteriol. **169**:3932–3937.
107. **Plamann, L. S., M. L. Urbanowski, and G. V. Stauffer.** 1988. *Salmonella typhimurium metE* operator-constitutive mutations. Gene **73**:201–208.
108. **Plumbridge, J. A.** 1990. Induction of the *nag* regulon of *Escherichia coli* by *N*-acetylglucosamine and glucosamine: role of the cyclic AMP-catabolite activator protein complex in expression of the regulon. J. Bacteriol. **172**:2728–2735.
109. **Polayes, D. A., P. W. Rice, M. M. Garner, and J. E. Dahlberg.** 1988. Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the *spf* gene of *Escherichia coli*. J. Bacteriol. **170**:3110–3114.
110. **Popham, D. L., D. Szeto, J. Keener, and S. Kustu.** 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. Science **243**:629–635.
111. **Ptashne, M.** 1986. A genetic switch: gene control and phage lambda. Cell Press (Cambridge, Mass.) and Blackwell Scientific Publications (Oxford, United Kingdom).
112. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.
113. **Queen, C., and M. Rosenberg.** 1981. A promoter of pBR322 activated by cAMP receptor protein. Nucleic Acids Res. **9**:3365–3377.
114. **Raibaud, O.** 1989. Nucleoprotein structures at positively regulated bacterial promoters: homology with replication origins and some hypotheses on the quaternary structure of the

activator proteins in these complexes. Mol. Microbiol. **3**:455–458.
115. **Raibaud, O., D. Vidal-Ingigliardi, and E. Richet.** 1989. A complex nucleoprotein structure involved in activation of transcription of two divergent *Escherichia coli* promoters. J. Mol. Biol. **205**:471–485.
116. **Rampersaud, A., S. Norioka, and M. Inouye.** 1989. Characterization of OmpR binding sequences in the upstream region of the *ompF* promoter essential for transcriptional activation. J. Biol. Chem. **264**:18693–18700.
117. **Ray, L., F. Claverie-Martin, P. Weglenski, and B. Magasanik.** 1990. Role of the promoter in activation of transcription by nitrogen regulator I phosphate in *Escherichia coli*. J. Bacteriol. **172**:818–823.
118. **Reitzer, L. J., R. Bueno, W. D. Cheng, S. A. Abrams, D. M. Rothstein, T. P. Hunt, B. Tyler, and B. Magasanik.** 1987. Mutations that create new promoters suppress the σ^{54} dependence of *glnA* transcription in *Escherichia coli*. J. Bacteriol. **169**:4279–4284.
119. **Reitzer, L. J., and B. Magasanik.** 1985. Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. Proc. Natl. Acad. Sci. USA **82**:1979–1983.
120. **Reitzer, L. J., and B. Magasanik.** 1986. Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. Cell **45**:785–792.
121. **Ren, Y. L., S. Garges, S. Adhya, and J. S. Krakow.** 1988. Cooperative DNA binding of heterologous proteins: evidence for contact between the cyclic AMP receptor protein and RNA polymerase. Proc. Natl. Acad. Sci. USA **85**:4138–4142.
122. **Rolfes, R. J., and H. Zalkin.** 1988. *Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator. J. Biol. Chem. **263**:19653–19661.
123. **Rolfes, R. J., and H. Zalkin.** 1988. Regulation of *Escherichia coli purF*. Mutations that define the promoter, operator, and purine repressor gene. J. Biol. Chem. **263**:19649–19652.
124. **Rolfes, R. J., and H. Zalkin.** 1990. Autoregulation of *Escherichia coli purR* requires two control sites downstream of the promoter. J. Bacteriol. **172**:5758–5766.
125. **Roovers, M., D. Charlier, A. Feller, D. Gigot, F. Holemans, W. Lissens, A. Piérard, and N. Glansdorff.** 1988. *carP*, a novel gene regulating the transcription of the carbamoylphosphate synthetase operon of *Escherichia coli*. J. Mol. Biol. **204**:857–865.
126. **Sakumi K., and M. Sekiguchi.** 1989. Regulation of expression of the *ada* gene controlling the adaptive response. Interactions with the *ada* promoter of the Ada protein and RNA polymerase. J. Mol. Biol. **205**:373–385.
127. **Sancar, A., G. B. Sancar, and W. D. Rupp.** 1982. LexA protein inhibits transcription of the *E. coli uvrA* gene *in vitro*. Nature (London) **298**:96–98.
128. **Sancar, G. B., A. Sancar, J. W. Little, and W. D. Rupp.** 1982. The *uvrB* gene of *Escherichia coli* has both *lexA*-repressed and *lexA*-independent promoters. Cell **28**:523–530.
129. **Santero, E., T. Hoover, J. Keener, and S. Kustu.** 1989. *In vitro* activity of the nitrogen fixation regulatory protein NIFA. Proc. Natl. Acad. Sci. USA **86**:7346–7350.
130. **Sasse-Dwight, S., and J. D. Gralla.** 1988. Probing co-operative DNA-binding *in vivo*: the *lac* O1:O3 interaction. J. Mol. Biol. **202**:107–119.
131. **Sasse-Dwight, S., and J. D. Gralla.** 1988. Probing the *Escherichia coli glnALG* upstream activation mechanism *in vivo*. Proc. Natl. Acad. Sci. USA **85**:8934–8938.
132. **Sasse-Dwight, S., and J. D. Gralla.** 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor Sigma 54. Cell **62**:945–954.
133. **Sawers, G., and A. Böck.** 1989. Novel transcriptional control of the pyruvate formate-lyase gene: upstream regulatory sequences and multiple promoters regulate anaerobic expression. J. Bacteriol. **171**:2485–2498.
134. **Schleif, R.** 1987. Gene regulation: why should DNA loop? Nature (London) **327**:369–370.
135. **Schleif, R.** 1989. Alternative loops and contacts by AraC protein, p. 25–30. *In* J. D. Gralla (ed.), DNA-protein interactions in transcription. Alan R. Liss, Inc., New York.
136. **Schlensog, V., and A. Böck.** 1990. Identification and sequence analysis of the gene encoding the transcriptional activator of the formate hydrogenlyase system of *Escherichia coli*. Mol. Microbiol. **4**:1319–1327.
137. **Schmitz, G., K. Nikaido, and G. F.-L. Ames.** 1988. Regulation of a transport operon promoter in *Salmonella typhimurium*: identification of sites essential for nitrogen regulation. Mol. Gen. Genet. **215**:107–117.
138. **Schwartz, M.** 1987. The maltose regulon, p. 1482–1502. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
139. **Shirabe, K., Y. Ebina, T. Miki, T. Nakazawa, and A. Nakazawa.** 1985. Positive regulation of the colicin E1 gene by cyclic AMP and cyclic AMP receptor protein. Nucleic Acids Res. **13**:4687–4698.
140. **Smith, A. A., R. C. Greene, T. W. Kirby, and B. R. Hindenach.** 1985. Isolation and characterization of the product of the methionine-regulatory gene *metJ* of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **82**:6104–6108.
141. **Smith, M. J., and H. A. Daum III.** 1986. Nucleotide sequence of the *purM* gene encoding 5′-phosphoribosyl-5-aminoimidazole synthetase of *Escherichia coli* K12. J. Biol. Chem. **261**: 10632–10636.
142. **Søgaard-Andersen, L., J. Martinussen, N. E. Møllegaard, S. R. Douthwaite, and P. Valentin-Hansen.** 1990. The CytR repressor antagonizes cyclic AMP-cyclic AMP receptor protein activation of the *deoCp2* promoter of *Escherichia coli* K-12. J. Bacteriol. **172**:5706–5713.
143. **Søgaard-Andersen, L., N. E. Møllegaard, J. R. Douthwaite, and P. Valentin-Hansen.** 1990. Tandem DNA-bound cAMP-CRP complexes are required for transcriptional repression of the *deoP2* promoter by the CytR repressor in *E. coli*. Mol. Microbiol. **4**:1595–1601.
144. **Stauffer, G. V., and L. T. Stauffer.** 1988. *Salmonella typhimurium* LT2 *metF* operator mutations. Mol. Gen. Genet. **214**:32–36.
145. **Stoner, C., and R. Schleif.** 1983. The *araE* low affinity L-arabinose transport promoter. Cloning, sequence, transcription start site and DNA binding sites of regulatory proteins. J. Mol. Biol. **171**:369–381.
146. **Storz, G., L. A. Tartaglia, and B. N. Ames.** 1990. Transcriptional regulator of oxidative stress-inducible genes: direct activation of oxidation. Science **248**:189–194.
147. **Straney, S. B., and D. M. Crothers.** 1987. Lac repressor is a transient gene-activating protein. Cell **51**:699–707.
148. **Su, W., S. Porter, S. Kustu, and H. Echols.** 1990. DNA-looping and enhancer activity: association between DNA-bound NTRC activator and RNA polymerase at the bacterial *glnA* promoter. Proc. Natl. Acad. Sci. USA **87**:5504–5508.
149. **Tartaglia, L. A., G. Storz, and B. N. Ames.** 1989. Identification and molecular analysis of *oxyR*-regulated promoters important for the bacterial adaptation to oxidative stress. J. Mol. Biol. **210**:709–719.
150. **Tobin, J. F., and R. F. Schleif.** 1987. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. **196**:789–799.
151. **Tobin, J. F., and R. F. Schleif.** 1990. Purification and properties of RhaR, the positive regulatory of the L-rhamnose operons of *Escherichia coli*. J. Mol. Biol. **211**:75–89.
152. **Tobin, J. F., and R. F. Schleif.** 1990. Transcription from the *rha* operon p(sr) promoter. J. Mol. Biol. **211**:1–4.
153. **Tommassen, J., M. Koster, and P. Overduin.** 1987. Molecular analysis of the promoter region of the *Escherichia coli* K-12 *phoE* gene: identification of an element, upstream from the promoter, required for efficient expression of PhoE protein. J. Mol. Biol. **198**:633–641.
154. **Tsung, K., R. E. Brissette, and M. Inouye.** 1989. Identification

of the DNA-binding domain of the OmpR protein required for transcriptional activation of the *ompF* and *ompC* genes of *Escherichia coli* by *in vivo* DNA footprinting. J. Biol. Chem. **264**:10104–10109.

155. **Tsung, K., R. E. Brissette, and M. Inouye.** 1990. Enhancement of RNA polymerase binding to promoters by a transcriptional activator, OmpR, in *Escherichia coli*: its positive and negative effects on transcription. Proc. Natl. Acad. Sci. USA **87**:5940–5944.
156. **Tuggle, C. K., and J. A. Fuchs.** 1986. Regulation of the operon encoding ribonucleotide reductase in *Escherichia coli*: evidence for both positive and negative control. EMBO J. **5**:1077–1085.
157. **Tuggle, C. K., and J. A. Fuchs.** 1990. Regulation of the operon encoding ribonucleotide reductase: role of the negative sites in *nrd* repression. J. Bacteriol. **172**:1711–1718.
158. **Ueno-Nishio, S., S. Mango, L. J. Reitzer, and B. Magasanik.** 1984. Identification and regulation of the *glnG* operator-promoter of the complex *glnALG* operon of *Escherichia coli*. J. Bacteriol. **160**:379–384.
159. **Urbanowski, M. L., L. S. Plamann, and G. V. Stauffer.** 1987. Mutations affecting the regulation of the *metB* gene of *Salmonella typhimurium* LT2. J. Bacteriol. **169**:126–130.
160. **Urbanowski, M. L., and G. V. Stauffer.** 1988. The control region of the *metH* gene of *Salmonella typhimurium* LT2: an atypical *met* promoter. Gene **73**:193–200.
161. **Urbanowski, M. L., and G. V. Stauffer.** 1989. Genetic and biochemical analysis of the MetR activator-binding site in the *metE metR* control region of *Salmonella typhimurium*. J. Bacteriol. **171**:5620–5629.
162. **Valentin-Hansen, P.** 1982. Tandem CRP binding sites in the *deo* operon of *Escherichia coli* K-12. EMBO J. **1**:1049–1054.
163. **Valentin-Hansen, P.** 1985. DNA sequences involved in expression and regulation of *deoR*-, *cytR*- and cAMP/CRP-controlled genes in *Escherichia coli*, p. 273–288. *In* R. E. Glass and J. Spizek (ed.), Gene manipulation and expression. Croom-Helm Ltd., London.
164. **Valentin-Hansen, P., B. Albrechtsen, and J. E. Love-Larse.** 1986. DNA-protein recognition: demonstration of three genetically separated operator elements that are required for repression of the *Escherichia coli deoCABD* promoters by the DeoR repressor. EMBO J. **5**:2015–2021.
165. **Valentin-Hansen, P., B. Holst, J. Josephsen, K. Hammer, and B. Albrechtsen.** 1989. CRP/cAMP- and CytR-regulated promoters in *Escherichia coli* K12: the *cdd* promoter. Mol. Microbiol. **3**:1385–1390.
166. **van den Elzen, P. J. M., J. Maat, H. H. B. Walters, E. Veltkamp, and H. J. J. Nijkamp.** 1982. The nucleotide sequence of the bacteriocin promoters of plasmids Clo DF13 and Col E1: role of LexA repressor and cAMP in the regulation of promoter activity. Nucleic Acids Res. **10**:1913–1928.
167. **Wang, Q., and J. M. Kagun.** 1989. dnaA protein regulates transcription of the *rpoH* gene of *Escherichia coli*. J. Biol. Chem. **264**:7338–7344.
168. **Watanabe, W., G.-I. Sampei, A. Aiba, and K. Mizobuchi.** 1989. Identification and sequence analysis of *Escherichia coli purE* and *purK* genes encoding 5′-phosphoribosyl-5-amino-4-imidazole carboxylase for de novo purine biosynthesis. J. Bacteriol. **171**:198–204.
169. **Webster, C., K. Gaston, and S. Busby.** 1988. Transcription from the *Escherichia coli melR* promoter is dependent on the cyclic AMP receptor protein. Gene **68**:297–305.
170. **Wek, R. C., and G. W. Hatfield.** 1988. Transcriptional activation at adjacent operators in the divergent-overlapping *ilvY* and *ilvC* promoters of *Escherichia coli*. J. Mol. Biol. **203**:643–663.
171. **Wilson, H. R., and C. L. Turnbough, Jr.** 1990. Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12. J. Bacteriol. **172**:3208–3213.
172. **Workman, J. L., and R. G. Roeder.** 1987. Binding of transcription factor TFIID to the major late promoter during *in vitro* nucleosome assembly potentiates subsequent initiation by RNA polI. Cell **51**:613–622.
173. **Yang, J., and. J. Pittard.** 1987. Molecular analysis of the regulatory region of the *Escherichia coli* K-12 *tyrB* gene. J. Bacteriol. **169**:4710–4715.
174. **Ye, S., and T. J. Larson.** 1988. Structures of the promoter and operator of the *glpD* gene encoding aerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. J. Bacteriol. **170**:4209–4215.
175. **Zhang, L., and J. D. Gralla.** 1989. Micrococcal nuclease as a probe for bound and distorted DNA in *lac* transcription and repression complexes. Nucleic Acids Res. **17**:5017–5028.